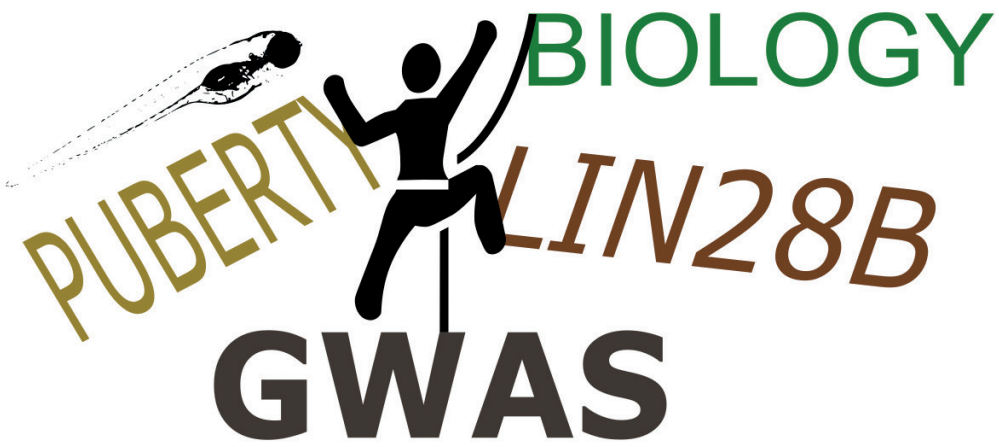
DISSERTATIONES SCHOLAE DOCTORALIS AD SANITATEM INVESTIGANDAM
UNIVERSITATIS HELSINKIENSIS

JAAKKO LEINONEN

MOVING BEYOND GWAS: EXPLORING THE FUNCTION OF THE GENE *LIN28B* ASSOCIATED WITH PUBERTAL TIMING



INSTITUTE FOR MOLECULAR MEDICINE FINLAND (FIMM)
FACULTY OF MEDICINE
DOCTORAL PROGRAMME IN BIOMEDICINE
UNIVERSITY OF HELSINKI

MOVING BEYOND GWAS: EXPLORING THE FUNCTION OF THE GENE *LIN28B* ASSOCIATED WITH PUBERTAL TIMING

Jaakko T. Leinonen

Institute for Molecular Medicine Finland, (FIMM)

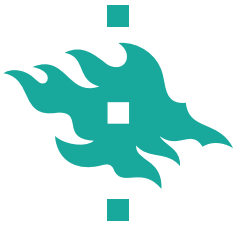
University of Helsinki, Helsinki, Finland

Doctoral Program in Biomedicine (DPBM)

ACADEMIC DISSERTATION

To be presented for public examination with the permission of the Faculty of Medicine of the University of Helsinki in Lecture Hall 2, Haartman Institute, Haartmaninkatu 3, on Thursday 10th of October 2019, at 11.00.

Helsinki 2019



ISBN - 978-951-51-5520-7 (paperback)
ISBN - 978-951-51-5521-4 (PDF)
ISSN 2342-3161 (print)
ISSN 2342-317X (online)
http://ethesis.helsinki.fi

Hansaprint Oy, Turku, Finland 2019

The Faculty of Medicine uses the Urkund system (plagiarism recognition) to examine all doctoral dissertations

Supervisor: **Elisabeth Widén**, MD, PhD
Finnish Institute for Molecular Medicine (FIMM)
Technology Centre
Centre of Excellence in Complex Disease Genetics
University of Helsinki, Finland

Reviewers: **Craig Primmer,** PhD, Academy Professor
Institute of Biotechnology
Evolution, Conservation, and Genomics
Organismal and Evolutionary Biology Research Programme
Helsinki Institute of Sustainability Science (HELSUS)
University of Helsinki, Finland

Reetta Hinttala, PhD, Associate Professor (Tenure track)
PEDEGO Research Unit
Faculty of Medicine
University of Oulu, Finland

Opponent: **Alexandre Reymond**, PhD, Professor
Faculty of Biology and Medicine
Center for Integrative Genomics
Université de Lausanne, Switzerland

Let's go.

Table of Contents

Abstract

Over the past decade, advances in genetics have led to identification of thousands of genetic loci in the human genome that contribute to complex traits such as the timing of pubertal onset. Following genome-wide association studies (GWAS), we currently know how a large proportion of the heritability of complex traits can be explained by common genetic variation in these loci. However, in many instances, we understand relatively little about the biology behind these genetic associations, and functional characterization of genetic variation has become a new bottleneck for genetic research. This thesis project was essentially motivated by the need to move beyond the GWAS associations. Prior to the project, sequence variants nearby *lin-28 homolog B* (*LIN28B)* had become associated with pubertal timing in the general population. Remarkably, compared to all other common genetic variants, the variants in the *LIN28B* locus appeared to exert relatively large effects on pubertal timing: one pubertal timing advancing allele associating with ~1,5 months advancement in age at menarche (AAM) in females. Yet , the gene, encoding for an RNA binding protein regulating cell division by binding *let-7* micro RNAs, showed little evidence of affecting the established mechanisms behind pubertal onset. Therefore, after our research group had been involved in establishing *LIN28B* as a pubertal timing gene, exploring the molecular mechanisms by which *LIN28B* affects puberty then became the focus of my thesis.

This thesis is founded on three original studies that have sought to elucidate *LIN28B* function from different perspectives. The first study utilized Finnish population cohorts to assess the potential pleiotropy of the gene in terms of body size and adult health. The second study combined human gene expression data from the GTEx database with zebrafish models, evaluating the consequences of transient dysregulation of *lin28b* during embryogenesis. For the third study, we first created *lin28b* knockout zebrafish with CRISPR-Cas9 technology to evaluate the effects of permanent *lin28b* knockout, and utilized the GTEx data and the UK biobank resource to study whether *LIN28B* contributes to sex steroid signaling in humans. Whereas the first study aimed to get an overview of the effects *LIN28B* has on human anthropometry and health, the second and the third were specifically aimed to elucidate the molecular mechanisms by which *LIN28B* affects traits like pubertal timing. In particular, this thesis project has focused on studying how *LIN28B* contributes to vertebrate development, and affects the function of the hypothalamic-pituitary (HP) axis.

The results presented in this thesis suggest that *LIN28B* has pleiotropic actions on vertebrate phenotypes and may contribute for example to the regulation of human body size in more complex ways than originally thought. Our studies demonstrate how the gene associates with variation in several body size parameters including hip size in humans, although showing little evidence of affecting metabolism in adults. The results also suggest that the effects that *LIN28B* has on body size are evolutionarily conserved. Based on models including transient manipulation of *lin28b* in zebrafish embryos, overexpression of the gene during embryogenesis appears to stimulate zebrafish growth, and intriguingly *lin28b* knockout zebrafish show similar growth patterns as humans that carry sequence variants linked with lower *LIN28B* expression. Importantly, the data presented in this thesis indicates that the sequence variants associating with pubertal timing affect *LIN28B* expression mostly in the hypothalamus and the pituitary of adult humans. It appears that this may have functional consequences: *LIN28B* expression at the HP axis seems to correlate positively with the expression of several hormonal genes like *ESR1* and *POMC*. Highlighting the gene's potential to contribute to sex steroid signaling, we finally associate *LIN28B* with regulation of testosterone levels in adult humans, which might be relevant in terms of explaining many of the GWAS associations. Overall, the original studies on which this thesis is based on have offered novel insight into *LIN28B* function in pubertal timing, heath and disease, enforcing the concept of the gene as a fundamental player in vertebrate growth and development. Besides helping to elucidate the function of *LIN28B*, the project has simultaneously served as an example for follow-up studies of GWAS loci in general.

Tiivistelmä

Edistysaskeleet genetiikan tutkimuksessa viimeisen kymmen vuoden aikana ovat johtaneet satojen sellaisten perimän kohtien tunnistamiseen, joissa esiintyvä vaihtelu vaikuttaa monitekijäisiin tauteihin ja ominaisuuksiin. Genominlaajuisten assosiaatiotutkimusten (GWAS) ansiosta tiedämme nykyisin kuinka merkittävä osuus monitekijäisten ominaisuuksien periytyvyydestä selittyy näiden kohtien yhteisvaikutuksella. Biologia näiden geneettisten assosiaatioiden taustalla on kuitenkin valtaosin huonosti ymmärrettyä. Tarpeesta ymmärtää miten nämä perimän kohdat vaikuttavat eri ominaisuuksiin ja tauteihin onkin muodostunut uusi pullonkaula geneettiselle tutkimukselle. Tämä väitöskirja sai alkunsa halusta ymmärtää murrosikään liittyvien GWAS-löydösten taustaa. Hieman ennen väitöskirjaprojektin alkua, ryhmämme oli ollut mukana liittämässä perimän vaihtelua *LIN28B*-geenin alueella murrosiän ajoittumiseen. *LIN28B* valikoitui tämän tutkimuksen kohteeksi erityisesti kahdesta syystä. Ensiksi, geenin lähellä sijaitseva geneettinen vaihtelu näyttää vaikuttavan murrosiän ajoittumiseen enemmän kuin mikään toinen yleisesti jaettu muutos genomissa: esimerkiksi naisilla yksi puberteettia varhaistava alleeli assosioituu n. 1,5 kuukautta varhaisempaan kuukautisten alkamiseen. Toiseksi, geeni ei kuitenkaan näyttänyt kytkeytyvän niihin mekanismeihin, joiden on aikaisemmin osoitettu vaikuttavan murrosiän alkamisen taustalla: *LIN28B*:n tiedettiin lähinnä kontrolloivan *let-7* mikroRNA-molekyylien biogeneesiä joka puolestaan oli liitetty solujen jakautumiseen ja erilaistumiseen. Väitöstyössäni keskityinkin selvittämään sitä, miten *LIN28B* voisi vaikuttaa murrosiän ajoittumiseen.

Tämä väitöskirja perustuu kolmeen osatyöhön, joissa on pyritty selvittämään LIN28B toimintaa eri näkökulmista. Ensimmäisessä tutkimuksessa hyödynsimme suomalaisia väestöaineistoja FINRISKI-tutkimuksista selvittääksemme voisiko geeni liittyä laajemminkin kehon koon säätelyyn ja aikuisterveyteen. Toisessa osatyössä yhdistimme geenilekspressiodataa GTEx tietokannasta seeprakalamallien kanssa, tutkiaksemme muun muassa miten GWAS tutkimuksissa löydetyt geenivariantit vaikuttivat *LIN28B*:n ilmentymiseen, ja kuinka sikiöaikainen muuntelu *lin28b*:n ekspressiomäärässä vaikuttaa kehitykseen. Kolmannessa osatyössä jatkoimme GTEx datan analyysiä ja loimme poistogeenisiä *lin28b*-seeprakalamalleja tutkiaksemme pysyvän *lin28b*:n hiljentämisen

vaikutuksia. Siinä missä ensimmäinen työ tarjosi yleiskatsauksen siihen kuinka *LIN28B* vaikuttaa kehon kokoon ja terveyteen liittyvään aineenvaihduntaan, toinen ja kolmas oli suunnattu erityisesti niiden molekyylitason mekanismien tutkimiseen joilla *LIN28B* voisi vaikuttaa esim. murrosiän ajoittumiseen. Näissä tutkimuksissa erityisenä kiinnostuksen kohteena olivat *LIN28B*:n mahdolliset vaikutukset hypotalamuksen ja aivolisäkkeen toimintaan.

Tässä väitöskirjassa esitetyt tulokset osoittavat *LIN28B*:n vaikuttavan selkärankaisten kasvuun ja kehitykseen monin eri tavoin: esimerkiksi ihmisen kehon koon muotoutumiseen monimutkaisemmilla tavoilla kuin aikaisemmin oli tunnettu. Kuitenkin, tutkimusten perusteella voitiin todeta, että geeni ei todennäköisesti vaikuta aineenvaihduntaan merkitsevästi aikuisväestössä. Väitöskirjassa esitetyt tulokset tukevat käsitystä *LIN28B* evoluutiossa säilyneestä roolista selkärankaisten kasvun säätelyssä. Hetkellinen *lin28B*:n yliekpressio seeprakalan alkioissa näyttäisi stimuloivan kalojen kasvua, siinä missä pysyvä *lin28b*:n hiljentäminen johtaa samanlaisiin muutoksiin kasvussa kuin ihmisillä joilla on matalampaan *LIN28B* ekspressiotasoon liittyvä muoto geenistä. Tutkimuksemme osoittavat lisäksi kuinka sama perimän vaihtelu joka liittyy murrosiän ajoittumiseen *LIN28B* geenin lähellä, vaikuttaa *LIN28B*:n ilmentymiseen erityisesti etenkin murrosiän kannalta tärkeissä kudoksissa, hypotalamuksessa ja aivolisäkkeessä. Tällä on mahdollisesti myös toiminnallisia seurauksia: *LIN28B* ilmentyminen näyttäisi korreloivan useiden hormonaalisten geenien ilmentymisen kanssa hypotalamuksessa ja aivolisäkkeessä. Korostaen geenin mahdollista roolia sukupuolihormoneihin liittyvään signalointiin liittyen, väitöstutkimukseni nostaa esiin erityisesti *ESR1*- ja *POMC*-geenit, sekä linkittää *LIN28B*:n testosteronitasojen säätelyyn aikuisilla, joka voisi selittää osan *LIN28B*:hen liitetyistä fenotyypeistä. Kaiken kaikkiaan, tutkimusprojektit jotka nyt esitellään kootusti tässä väitöskirjassa ovat siis tuoneet lisätietoa siihen miten *LIN28B* vaikuttaa murrosikään, terveyteen ja yksilönkehitykseen, vahvistaen käsitystä geenistä merkittävänä tekijänä selkärankaisten kasvun säätelyssä. Paitsi että väitöstutkimukseni on tuonut uutta tietoa niihin molekyylitason mekanismeihin joilla *LIN28B* mahdollisesti vaikuttaa kasvuun ja kehitykseen, projektimme toimii myös yleisellä tasolla esimerkkinä keinoista joilla voidaan lähteä tutkimaan GWAS-löydösten biologista taustaa.

Original Publications

This thesis is based on the following original publications, which are referred to in the text by their Roman numerals:

I. **Leinonen JT**, Surakka I, Havulinna AS, Kettunen J, Luoto R, Salomaa V, Widén E. Association of *LIN28B* with Adult Adiposity-Related Traits in Females. PLoS One 2012;7(11):e48785

II. **Leinonen JT**, Chen YC, Tukiainen T, Panula P, Widén E. Transient manipulation of *lin28b* expression – permanent effects on zebrafish growth. Mol Cell Endocrinol. 2019 Jan 5;479:61-70

III. **Leinonen JT**, Chen YC, Pennonen J, Junna N, Tukiainen T, Panula P, Widén E. *lin28b* regulates zebrafish growth affecting gene expression at the hypothalamic-pituitary axis. *Submitted*.

Abbreviations

2D4D	second digit -fourth digit (fore finger -ring finger)
AAM	age at menarche
ACTH	adrenocorticotrophic hormone
BMD	bone mineral density
BMI	body mass index
CCHC	CysteinCysteinHistidineCystein
CPM	counts per million
CPP	central precocious puberty
CRISPR	clustered regularly interspaced short palindromic repeats
CSC	cancer stem cells
CSD	cold shock domain
DIG	Digoxigenin
DLK1	Delta-like 1 homolog
DPF	days-post-fertilisation
DSB	double stranded breaks
ENU	N-ethyl-N-nitrosurea
eQTL	expression quantitative trait locus
ESR1	estrogen receptor 1
ESRα	estrogen receptor alpha
ExAC	Exome Aggregation Consortium
FinnGen	Finnish Genomes
FSH	follicle-stimulating hormone
GABA	gamma-aminobutyric acid
GH	growth hormone
GnomAD	Genome Aggregation Database
GnRH	Gonadotropin-releasing hormone
GNRHR	gonadotropin-releasing hormone receptor
GO	gene ontology
GORILLA	Gene Ontology enRIchment anaLysis and visuaLizAtion
gRNA	guide RNA
GSEA	Gene Set Enrichment Analysis
GTEx	Genotype-Tissue Expression
GWAS	Genome-wide association study
HDL	high-density lipoprotein
HH	hypogonadotrophic hypogonadism
HP	hypothalamic-pituitary
HPF	hours-post-fertilisation
HPG	Hypothalamic-pituitary-gonadal

HRM	high resolution melting curve
iPS	induced pluripotent cells
KISS1	Kisspeptin
KISS1R	kisspeptin receptor
KO	Knockout
let-7	lethal 7
LH	lutenizing hormone
lin-28	lineage 28
MAF	minor allele frequency
miRNA	micro RNA
MKRN3	makorin RING finger protein 3
MO	morpholino oligonucleotide
NMDA	N-methyl-D-aspartate
ORF	open reading frame
POMC	Proopiomelanocortin
REViGO	REduce + Visualize Gene Ontology
RIN	RNA Integrity Number
RPKM	reads per kilobase million
SNP	single nucleotide polymorphism
T2D	Type 2 Diabetes
TAC3R	tachykinin receptor 3
TILLING	Target Induced Local Lesion in Genomes
UKBB	UK Biobank
WES	whole exome sequencing
WGS	whole genome sequencing
WHR	waist-to-hip ratio
ZFN	zinc finger nuclease

Introduction: Pubertal timing, a crucial event in development - *LIN28B*, a crucial gene for pubertal timing

The onset of puberty is a crucial milestone in development, marking the beginning of sexual maturation, the key developmental stage between childhood and adulthood. As a developmental event, the onset of puberty is distinctive in nature, having dramatic consequences for both the appearance and behavior in all vertebrates. Nonetheless, instead of being an isolated event in life, pubertal onset can be perceived as a culmination point of all the prior development that has taken place to ensure sexual maturation can proceed. This development begins at the time of conception, terminating at the second decade of human life. On average, the timing of pubertal onset in humans varies substantially between individuals, sexes and populations, usually occurring between the ages of 8 and 14 in modern societies (1-3). Importantly, this essential milestone in human development, having strong genetic basis, is known to correlate with several body size parameters and risk for many diseases (4-6).

Despite the evidence that heritability of pubertal timing is relatively high (50-80%) and although several monogenic syndromes affecting pubertal development had been identified, relatively little was known about the contribution of individual genes to pubertal timing in the healthy population when the work for this thesis begun in 2011 (6-8). At the same time, technological and analytical advances had made it possible to conduct large-scale genome-wide association studies (GWAS) to elucidate the genetic background of puberty. Our research group had been involved in identifying *LIN28B* – the gene under the focus of this thesis - as one of the first genetic loci where common genetic variation could be linked with changes in pubertal timing and growth in the healthy population (9). In fact, the variants nearby *LIN28B* still appear to have the biggest effects on pubertal timing out of all examined common variants (10). Prior to GWAS, studies in the nematode *C. elegans* had already indicated that *LIN28*-genes might have potential to affect developmental timing (11). However, at the time of the initial discovery in humans, *LIN28B* showed no direct biological links to the established physiological mechanisms that regulate pubertal onset. Since then several GWAS, nowadays encompassing massive number of subjects, have identified hundreds of more loci contributing to pubertal timing in the general population (10, 12, 13).

Meanwhile, the *LIN28B* association with pubertal timing has remained robust, whilst the gene has become associated also with many other traits, ranging from depression and intraocular pressure to bone mineral density (BMD) and height (14-18). Yet, the mechanisms by which the gene affects all these phenotypes have remained somewhat elusive.

Essentially, the gap in the knowledge linking *LIN28B* with puberty provided the ultimate motivation for this thesis. In a wider context, the case of *LIN28B* and puberty however is not an unusual case: in the wake of the GWAS era, the bottleneck for genetic research is not anymore "finding the right gene" as it was in the past. Currently, larger and larger studies like the UK biobank study (UKBB) and the FinnGen project (www.finngen.fi/) have yielded, and will yield a huge number of replicable genetic associations with a vast number of traits (19). Instead, to find out the biological mechanisms by which the identified genetic loci contribute to various phenotypes remains as a big challenge (20). Providing an example of a project that aims to take the leap from the associations into understanding the underlying biology, this thesis has concentrated on unraveling the mechanisms behind the *LIN28B* associations with growth and pubertal timing. In the studies forming the body of this thesis, we have used several different approaches to "move beyond the GWAS results". To study *LIN28B* function, we have utilized data from Finnish population cohorts and international gene expression databases, and combined these with our own in-house animal models that reveal interesting aspects about the physiology of pubertal timing and the function of *LIN28B*.

The next chapters of this thesis aim to introduce the known physiology and hormonal mechanisms behind pubertal timing, with a special focus on the development of the hypothalamic-pituitary-gonadal (HPG) axis. Following description of what is currently known about the function of *LIN28B*, the literature review concentrates on the remaining question of how to move forward from the genetic associations into understanding the biology: the required step to fully translate the value of the scientifically solid genetic associations. After the literature review, results from the three original studies serving as the basis of this thesis are recapped and discussed. The final chapter then aims to summarize what we have learned about the mechanisms linking *LIN28B* with pubertal timing, and what we have learned about studying GWAS loci in general.

2. Literature review

2.1 The physiology of pubertal onset and epidemiology of pubertal timing

2.1.1 Setting the stage: embryogenesis and childhood development

In humans, the physical development from child to adult is a continuous process that takes almost two decades, culminating at pubertal period that marks the transition from childhood into sexual maturity. Traditionally, sexual maturation - the attainment of reproductive capability - has been considered as the final physical step between childhood and adulthood, although the development of the human brain and secondary sexual characteristics may continue for several years beyond this point (21). Importantly, past research has shown that the onset of puberty marks a huge shift in the secretion of sex steroids (androgens, estrogens and progestogens), causing the physiological changes associated with puberty (22, 23). The central event triggering the sex steroid secretion is the establishment of pulsatile release of gonadotropin-releasing hormone (GnRH) from the hypothalamus. This event is currently relatively well characterized, and many factors that affect GnRH secretion have been identified, including neuropeptides kisspeptin and dynorphin, and balance of glutamatergic vs. GABAergic stimulation of the GnRH neurons (24, 25). However, the physiological and environmental mechanisms that precede, and ultimately trigger the GnRH secretion and pubertal development remain poorly understood (24, 25).

Although the onset of puberty appears to be a distinctive event in human development, it can be seen as the final event in childhood growth that in turn has foundations in fetal growth, beginning upon fertilization. In fact, the tissues responsible for pubertal onset, forming the hypothalamic-pituitary-gonadal (HPG) axis, develop already during embryogenesis (26-28). The HPG axis consists of specific neurons and glands, including the hypothalamic GnRH neurons and the glands secreting the pituitary gonadotrophins (lutenizing hormone (LH) and follicle-stimulating hormone (FSH)) that stimulate sex hormone release from the gonads. Remarkably, once formed, the HPG axis shows transient activation

already before puberty, but becomes permanently activated only when the pubertal development begins (26-28). Studies on human fetuses have shown that HPG activity is first detected during midgestation (26-28). Towards the end of pregnancy, this activity declines, but is again transiently restored after birth (26-28). The biological significance of these periods of HPG activity before puberty is not completely understood, but it appears necessary for example for the development of the male genitalia, and likely induces changes also in the organization of neural networks that contribute to sex-specific behavior (26-28). Supporting the concept that the HPG axis forms early and is thereafter potent to induce gonadotrophin secretion upon suitable stimuli, animal experiments have shown that the axis can be permanently activated already during childhood. For example, in rhesus monkeys, spermatogenesis and ovulation can be induced prematurely by hormonal treatment (25). Interestingly, animal experiments and clinical observations in humans tell also the complementary story: various conditions causing low gonadotrophin levels, reflecting the malfunction of the HPG axis, can lead to delayed or absent puberty (25). In some cases these conditions are known to be caused by defects in fetal development. For example, the GnRH neurons are not hypothalamic in origin, but rather migrate to the hypothalamus from the olfactory placode during embryogenesis (29, 30). Failures in this migration, leading to Kallman syndrome, have been associated with problems in later pubertal development (31, 32). In the light of the current research, although the exact molecular mechanisms triggering pubertal onset are incompletely understood, puberty can anyhow be viewed as a unique event in human growth that is ultimately based on successful embryonic and childhood development.

2.1.2 The mechanisms of pubertal onset

The studies on the mechanisms responsible for pubertal onset truly took off at the second half of the 20th century, a couple of decades after sex hormones had been identified and isolated (25, 33, 34). The hypothesis that a specific neural stimulus that triggers pituitary gonadotrophin release activates pubertal development was finally confirmed when two independent research groups isolated GnRH, a hypothalamic decapeptide in 1971 (25, 35, 36). Since then, a lot of research has been concentrated on understanding the molecular mechanisms that are responsible for the pulsatile GnRH release associated with pubertal

onset. Naturally, both the development and function of the GnRH neurons has been studied extensively. However, the formation of the HPG axis is not limiting for pubertal onset, as the axis is ready to function already during the fetal stages, a lot of emphasis has been put on studying which factors are responsible for re-activating the system after the quiescent childhood period (24, 25, 37). The current theory holds that puberty begins when multiple different genetic and environmental signals converge and signal the brain to activate the HPG axis (**Figure 1**) (24, 25, 37, 38). The hallmark of this activation is the re-activation of the pulsatile GnRH release from the hypothalamus, causing gonadotrophin secretion from the pituitary. The theory predicts that there exists a type of neurobiological brake that has been on since late infancy, but is switched off to launch pubertal development in humans. This brake has been conceptualized to consist of several neurotransmitters, neuropeptides and environmental signals. Whether the brake is simply related to the somatic development, or is regulated by a clock-like mechanism in the brain has not been properly addressed (24, 25, 38). Similarly, the mechanisms that switch on the brake in the childhood remain unknown. Nonetheless, all research done over the past decades points towards the concept that there is no single biological signal or a gene that is responsible for pubertal onset: rather, we now known that there exists many "permissive" factors that seem to affect the initiation of pubertal development (38).

At present, it is thus widely accepted that integration of several neuronal signals to stimulate the pulsatile GnRH release upon puberty is required. In simple terms, this pulsatile GnRH release activity depends on the balance of excitatory and inhibitory inputs that originate either from hypothalamic neurons or glial cells (24, 25, 38). Recent genetic studies have identified several genes that seem to act as upstream regulators of the GnRH secretion. For example, genetic studies on hypogonadotrophic hypogonadism (HH), a syndrome leading to absent puberty, have highlighted the importance of kisspeptin (*KISS1*), neurokinin B and their receptors (*KISS1R* and *TAC3R*, respectively) for proper GnRH secretion (39-41). Correspondingly, functional studies have shown that both KISS1 and neurokinin B can stimulate GnRH release (24, 42, 43). Although these genes indeed participate to the cascade that triggers the GnRH release, it is worth noting that they do not directly "regulate" pubertal timing. Rather, these genes can be viewed as downstream players that are activated after the neurobiological brake has been switched off.

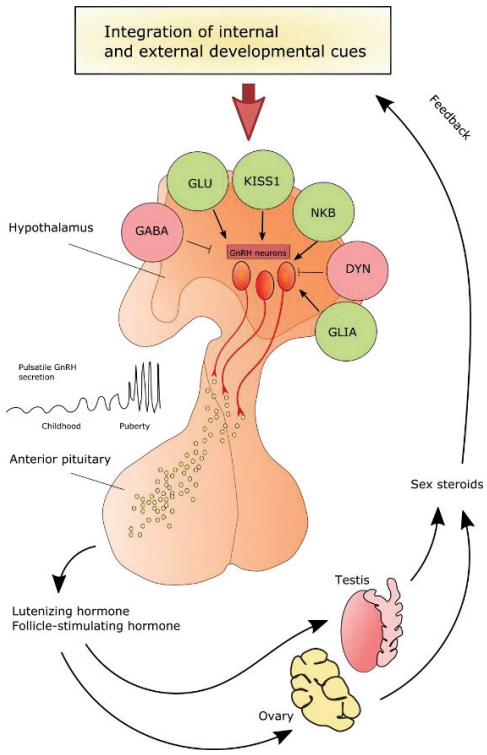


Figure 1. A schematic model of the function of the hypothalamic-pituitary-gonadal (HPG) axis that controls gonadotrophin-releasing hormone (GnRH) pulse generation triggering puberty. To launch puberty, poorly understood neural mechanisms activate Kisspeptin (KISS1) neurokinin B (NKB) and glutamatergic (GLU) neurons, leading to stimulation of GnRH pulse generation and GnRH release from the hypothalamus. Glial cells are also known to stimulate GnRH release. GABAergic neurons (GABA= γ-aminobutyric acid) and dynorphin (DYN) expression from the arcuate nucleus have been suggested to be part of a prepubertal brake on GNRH secretion. Ultimately, the pulsatile GnRH release leads to secretion of LH and FSH, which then stimulate gonads to secrete sex steroids. The sex steroids testosterone and estrogen are responsible for example of the growth spurt associated with pubertal timing, and the development of the secondary sexual characteristics. Figure based on Sisk and Foster 2004 & Herbison 2016(24, 38).

In combination with the genetic studies, physiological and "omics" studies have been successful in identifying molecular mechanisms and factors that are crucial for the regulation of the pubertal onset. Studies on the role of neurotransmitters in the regulation of puberty have linked changes in the γ-aminobutyric acid (GABA) and glutamate signaling with pubertal onset (44, 45). Notably, pubertal onset has been correlated with a decrease in GABA, the major inhibitory neurotransmitter in monkey brain, but whether GABA contributes directly to GnRH regulation remains poorly understood. Contrastingly, GnRH release can be stimulated in monkeys by glutamate analog N-methyl-D-aspartate (NMDA) (46). Other peptides and mechanisms related to pubertal timing include dynorphin, an opioid peptide which has been associated with inhibition of GnRH secretion (47). Lately, also a role for hypothalamic glial cells in stimulating pubertal onset has been proposed (37). Nevertheless, the same limitation applies to these findings as in the case of KISS1 and neurokinin B: although changes in the content of these neurotransmitters in critical regions of the brain

would somehow be related to triggering puberty, the ultimate cause for these changes remains unknown.

2.1.3 Pubertal timing , a complex trait showing normal distribution and sex differences in the general population

Although the basic physiological mechanisms that lead to pubertal onset have been characterized, the factors that affect the timing of puberty are less well understood. In epidemiological terms, pubertal timing represents a complex trait that varies roughly according to the normal distribution in the general population. Like other complex traits, pubertal timing is thought to be determined by the interaction of multiple genetic and environmental factors, which will be discussed in more detail in the following sections of this literature review. As a result of these factors, the timing of pubertal onset may vary considerably between given two individuals, although for most people the pubertal onset still occurs close to the population mean.

Like most physiological traits, pubertal timing as well as pubertal tempo (how fast the puberty proceeds) differs between sexes in humans. On average, girls start to show the first signs of pubertal development between 8 and 13 yrs of age, whereas the earliest manifestations of puberty can be usually seen in boys between the ages 9 and 14 (1-3). **Figure 2** shows illustration of the timing of pubertal development for both sexes. The fact that girls enter puberty earlier than boys is well established, but the reasons for this phenomenon remain somewhat unclear. Interestingly, it has been shown that girls tend to optimize brain connections earlier than boys (48). Therefore, it can be speculated that the sex differences in pubertal timing might be directly related to the maturation of the neural networks and hormonal mechanisms that ultimately control pubertal onset. Compared to most animals, humans exhibit relatively late onset of puberty. This type of sex difference in the maturation rate is still not specific to humans: for example in savanna baboons and several other animals males usually mature later than females (49). Notably, examples of the opposite pattern in other species also exist: for example male salmons tend to mature earlier and at smaller size than females (50).

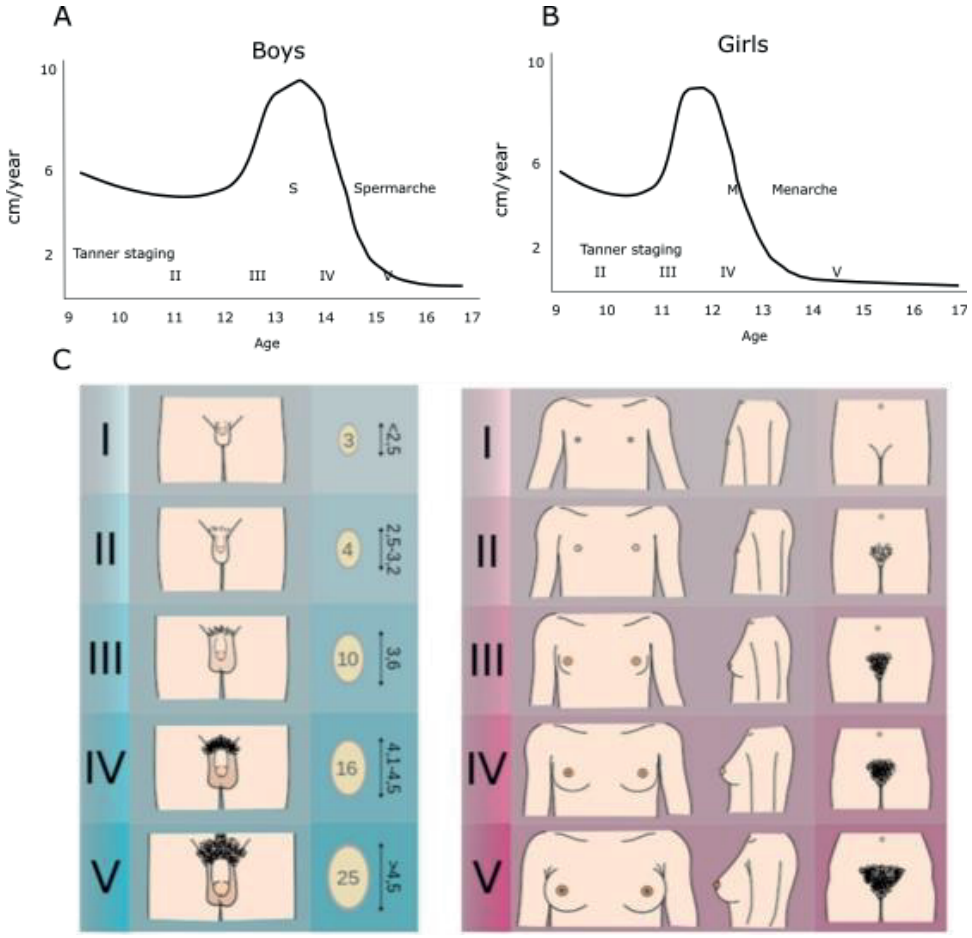


Figure 2. Schematic representation of timing of pubertal development in boys and girls. Panels A and B show mean growth in cm/year and rough age for peak height velocity, spermarche, menarche and Tanner stages for boys and girls. Panel C shows visualization of Tanner stages. Adapted and based on images from Physiologic Growth and Development During Adolescence David S. Rosen (2004) and Michal Kormorniczak (shared under creative commons license)(51).

2.1.4 Pubertal timing and links to body size

As a physiological phenomenon, pubertal timing is closely intertwined with the development of several anthropometric traits, and some well-known, though complex correlations between pubertal timing and body size in humans exist. Perhaps the most studied correlations exist between pubertal onset and obesity. In girls, earlier sexual maturation has been associated with higher body mass index (BMI), whereas physical activity and resulting lower body fat correspondingly associate with later pubertal onset (52-55). The observation that increased BMI in childhood leads to earlier pubertal onset in girls is likely related to the hypothesis that before entering puberty the brain needs to get signals from the body that enough growth has occurred to support the potential consequences of the sexual maturation, namely raising the offspring (52). Contrastingly, in boys higher BMI does not seem to advance pubertal maturation: in some studies obesity in boys has even been associated with delayed puberty (56). Noteworthy, however, the relationship between puberty and obesity in boys has been less studied compared to girls.

Whereas obesity may directly modulate pubertal timing, the links between puberty and height may be more complex, though especially the former is known to affect the latter. In girls, earlier pubertal onset seems to associate with shorter adult height, potentially due to subsequent earlier cessation of growth (57, 58). In boys, the results are again somewhat different: studies have suggested that earlier pubertal maturation may correlate even with slower pubertal progression and taller stature as adults, though this remains under debate (6, 57). Highlighting the complexity of growth and the interplay between height, obesity and puberty, increased height seems to correlate with childhood obesity, but this correlation disappears in adults (59, 60). It is generally accepted that part of these correlations might be explained by the actions of various hormones including growth hormone (GH), leptin and the sex hormones, but overall the developmental mechanisms contributing to body size and pubertal onset seem to be complex.

2.1.5 Disturbances in pubertal timing

Based on the normal distribution, for most people the pubertal onset occurs ±1 yrs from the population mean. Yet, pubertal onset that falls at the extreme tails of the normal distribution might be indicative of a medical condition (1, 3). Traditionally, three classes of pubertal timing disorders are recognized. Pubertal timing can happen either exceptionally early (precocious puberty) or late (delayed puberty), and in rare cases, the initiation of puberty might be completely stalled (absent puberty). Due to undefined reasons, it appears that girls are more likely to suffer from precocious puberty, whereas boys are more likely to show delays in puberty(61). Conventionally, the limits for abnormally precocious sexual maturation have been set as puberty initiation of less than 8 yrs for girls and 9 yrs for boys (1, 3, 61). On the other hand, puberty is considered to be delayed if a person does not show any apparent signs of pubertal development by the age of 13 (girls), or 14 (boys) (1-3). Notably, these limits are largely based on epidemiological studies made in the UK several decades ago, and thus not might directly apply to every modern population (62, 63).

2.1.6 Heritability of pubertal timing

Although the onset of puberty may vary extensively between individuals, generally the timing of puberty appears highly heritable. It has been estimated that ~50-80% of the variation in pubertal timing is affected by genetic factors in western societies (3, 6, 64). The evidence for such a high genetic contribution comes from the combination of many sources. For example, the average timing of pubertal onset is known to vary from population to population, and even within a population based on ethnic group (1, 64, 65). It is also known to show a high correlation between family members, and twin studies have indicated that pubertal timing varies less between monozygotic than dizygotic twins (66). The data estimating the heritability of pubertal timing is largely based on studying the age at menarche (AAM), but interestingly the analyses of other proxies of pubertal development such as the height growth spurt have estimated the genetic contribution to be even higher (>90%) (6). Like in the case of some other physical traits (e.g height), the estimated relatively high heritability might be a reflection of the fact that environmental influences such as disease and undernutrition contribute relatively little to pubertal timing in modern societies.

2.1.7 Environmental influences on pubertal timing

Despite the apparently high heritability of pubertal timing, environmental factors may still have a major influence on the onset of puberty, both at the population and individual level. For example the age references for normal pubertal timing still vary between well-off and underprivileged countries, largely due to environmental factors, including diet (1). On the other hand, the timing of puberty has changed in the western world over the past century. In many countries, a secular trend in pubertal timing has been evident, meaning that the average age when pubertal maturation starts has continue to decline. For example in Finland, the average AAM used to be >17yrs in the middle of the 19th century (67), whilst by 1969 it had come down to 13.3yrs (68), being around 13.3 yrs still in 1993(1). Similarly, in the USA the average AAM has fallen from 13.5 yrs to 12.3 yrs when comparing females born in 1908-9 to those born between the years 1990-93 (69).

The potential reasons behind these secular trends may vary. Yet, since the genetics of a given population are unlikely to change as rapidly as the average AAM has fallen, these are likely related to environmental changes. One of the key environmental contributions to growth is nutrition. Data from several sources indicates that maturation is largely dependent on nutritional conditions in many species ranging from invertebrates to humans (1, 49). In simple terms, this means that for puberty to begin, a critical amount of growth has to occur. For example zebrafish do no become sexually mature before reaching a threshold weight, and some studies have estimated that normal menstrual cycling in females requires at least 22% of body fat (1, 70). In many animals additional factors related to ecology and socio-demography are key contributors to the timing of sexual maturation, partly through affecting nutrient availability between individuals. Also in humans, factors like stress, climatic conditions, and endocrine-disrupting chemicals have been speculated to contribute to the variation and secular trends in pubertal timing (1).

2.1.8 Pubertal timing and health

Pubertal timing and growth patterns during early life have been associated with many adult health related outcomes. Both early and late pubertal timing have been shown to correlate significantly with disease risk (4). Particularly early pubertal timing has been robustly linked with all-cause mortality and the risk for several disease and traits relating to variable health conditions. These include for example cardiometabolic traits like higher adult BMI, fasting insulin, diastolic blood pressure, and decreased HDL cholesterol in both sexes, as well as disease like breast cancer and type 2 diabetes (4, 5, 71). As early onset of puberty has been shown to correlate with obesity, some of the associations might be directly related to BMI and adiposity. The role of obesity as a causal factor behind some of the associations has in fact been strengthened by many epidemiological and genetic studies, but pubertal onset seems to be also an independent risk factor for many diseases (4). A recent UKBB study including ~450,000 individuals showed how pubertal timing may directly affect the risk for several diseases, including some cancers, cardiovascular, musculoskeletal, gastrointestinal and neurological disorders in both sexes (4). A brief summary of these associations is presented in **Table 1**. Notably, both early (AAM between 9 and 11 yrs) and late (AAM between 15 and 17 yrs) puberty seem to affect adult health. Overall, the existing epidemiological associations between pubertal onset and health highlight the fact that ontogenesis and development of disease are closely related processes, and understanding the mechanisms that affect pubertal onset may have immediate medical relevance.

Table 1. List of selected traits that show epidemiological association either with early or late pubertal timing in the UKBB cohort. The traits that associate either with early or late pubertal timing range from cancer to poor overall health. Bolded *p*-values (P) refer to statistically significant correlations after correction for multiple testing. The "adjusted" refers to a statistical model that includes obesity and socioeconomic status as additional covariates. The adjusted odds ratios (OR) are generally lower, indicating that the epidemiological associations with pubertal timing are partly related to obesity. Table has been adopted from Day et al. "Puberty timing associated with diabetes, cardiovascular disease and also diverse health outcomes in men and women: the UK Biobank study" (4).

Early Menarche (8–11 years)					Late Menarche (15–19 years)					
	Unadjusted		Adjusted		Unadjusted		Adjusted		Linear Trend	
Disease	OR	P	OR	P	OR	P	OR	P	P	P adj.
Cancer										
Breast cancer	1.09 (1.03-1.16)	0.002	**1.13 (1.06-1.20)**	**5.6E-05**	0.94 (0.88-1.00)	0.04	0.94 (0.88-1.00)	0.05	**4.3E-08**	**9.2E-11**
Cardiovascular/Metabolic/Endocrine										
High cholesterol	**1.23 (1.18-1.28)**	**<1E-21**	1.06 (1.02-1.11)	0.006	**1.09 (1.05-1.14)**	**6.3E-05**	1.02 (0.98-1.07)	0.30	**2.3E-11**	0.009
Hypertension	**1.37 (1.33-1.41)**	**<1E-21**	**1.13 (1.1-1.16)**	**1.6E-15**	0.96 (0.93-0.99)	0.006	0.96 (0.93-0.99)	0.008	**<1E-21**	**<1E-21**
Obesity†	**1.91 (1.86-1.96)**	**<1E-21**	**1.82 (1.77-1.87)**	**<1E-21**	**0.85 (0.82-0.88)**	**<1E-21**	**0.76 (0.74-0.79)**	**<1E-21**	**<1E-21**	**<1E-21**
Type 2 diabetes	**1.76 (1.62-1.91)**	**<1E-21**	**1.25 (1.15-1.36)**	**5.2E-07**	0.94 (0.85-1.04)	0.22	0.90 (0.81-1.0)	0.05	**<1E-21**	**2.1E-11**
Gastrointestinal disorder										
Allergy to food	**1.41 (1.21-1.65)**	**1.0E-05**	**1.39 (1.19-1.62)**	**3.7E-05**	1.08 (0.91-1.28)	0.40	1.12 (0.95-1.34)	0.18	1.3E-03	0.02
Malabsorption/Coeliac disease	0.87 (0.73-1.03)	0.11	0.95 (0.8-1.13)	0.55	**1.69 (1.45-1.98)**	**1.9E-11**	**1.62 (1.39-1.89)**	**8.6E-10**	**<1E-21**	**4.3E-12**
Musculoskeletal disorder										
Arthritis	**1.35 (1.19-1.52)**	**2.5E-06**	1.20 (1.06-1.36)	0.005	0.99 (0.86-1.14)	0.87	0.94 (0.81-1.09)	0.41	**2.9E-06**	8.7E-04
Osteoporosis	**0.84 (0.78-0.91)**	**7.1E-06**	0.91 (0.84-0.98)	0.02	**1.17 (1.09-1.26)**	**2.9E-05**	1.09 (1.01-1.18)	0.03	**<1E-21**	**2.5E-06**
Short stature	**1.58 (1.50-1.67)**	**<1E-21**	**1.19 (1.09-1.30)**	**1.2E-05**	0.96 (0.91-1.02)	0.24	**0.80 (0.72-0.89)**	**2.4E-05**	**<1E-21**	**<1E-21**
Neuro/Cognition/Psychiatry										
Depression	**1.19 (1.14-1.25)**	**1.1E-13**	1.05 (1.01-1.10)	0.03	**1.12 (1.07-1.18)**	**6.3E-06**	1.07 (1.02-1.13)	0.008	0.007	0.36
Respiratory/ENT disorder										
Asthma	**1.17 (1.13-1.21)**	**<1E-21**	1.06 (1.03-1.1)	8.3E-04	**1.10 (1.06-1.14)**	**9.4E-07**	**1.11 (1.07-1.15)**	**7.9E-08**	1.6E-04	0.02
Other										
Poor overall health	**1.51 (1.42-1.60)**	**<1E-21**	1.10 (1.03-1.18)	0.003	**1.33 (1.25-1.43)**	**<1E-21**	**1.19 (1.11-1.27)**	**1.3E-06**	**8.2E-07**	0.08

2.2. The genetics of puberty

2.2.1 The genetics of puberty – a complex affair

Pubertal timing is a genetically complex, polygenic trait, whereas disorders causing either complete absence of puberty or severe cessation of pubertal development appear very rare in humans. For example the prevalence of HH, the most common collection of syndromes leading to absent puberty has been estimated to be around 1:10000-86000, including both congenital (genetic) and acquired cases in the US (72). Likewise, the prevalence of Kallman syndrome, a special form of HH is around 1:30000 in Finnish men, and 1:125000 in Finnish women (73).

Although the genetic factors that explain most of the normal variation in pubertal timing in human populations are likely separate from those that underlie severe syndromes affecting the pubertal onset, these syndromes have anyhow offered critical insight in to understanding the basic mechanisms that trigger the puberty. The clinical spectrum of these syndromes varies to some extent. These may result from defects related to the function of the gonads, preventing normal sex hormone secretion (hypogonadism), or from lesions directly affecting the GnRH system, in which case the condition is a called *hypogonadotrophic* hypogonadism (HH). The inheritance patterns for genetic pubertal timing disorders also vary: they may result from dominant *de novo* – mutations, show X-linked inheritance pattern, or be caused by recessive mutations that are often inherited in compound heterozygous form (8, 65). On top of the complex nature of pubertal genetics, epigenetic mechanisms have been shown to participate to regulation of the pubertal processes (74).

Currently, the knowledge about the genes that contribute to pubertal timing has increased almost exponentially, and several excellent reviews about the current state of pubertal genetics have been written (65, 75, 76). The next chapters aim to provide a brief overview on the genetics of pubertal timing, concentrating first on established monogenic syndromes,

followed by an overview of the recent advances in understanding the polygenic contribution to puberty in the general population.

2.2.2 Monogenic forms of pubertal timing disorders

The studies on heritability of pubertal timing have shown that genes have considerable potential to affect the age at pubertal onset. Like for many diseases and traits, the first genetic insights onto the mechanisms that are responsible for the pubertal onset came from family studies which concentrated on identifying genetic lesions that interfere with pubertal progress. These can be roughly divided into two categories: mutations that cause an early puberty, and mutations leading to pubertal progress being delayed or blocked. Several mutations that either advance or delay the onset of puberty have been currently identified. For example, mutations in more than 20 genes have been linked with syndromes that cause absent puberty (65, 75). Remarkably, emphasizing the pivotal mechanism behind pubertal onset, in most cases the mutations causing either precocious or absent puberty have been found to affect the hypothalamic GnRH release.

Perhaps the best examples of the well-characterized monogenic syndromes related to pubertal timing are those causing an absent puberty. Especially studies on congenital HH and its distinctive form, the Kallman syndrome (collectively referred to as isolated GnRH deficiencies), have identified many key genes that contribute to pubertal onset. In general, the genes linked to absent puberty are immediately related to GnRH function. Usually, mutations in these genes prevent gonadotrophin secretion from the pituitary by interfering with the normal development of the GnRH neurons and hence GnRH release in the hypothalamus. The HH generally results from mutations in hypothalamic genes that participate to the GnRH release, whereas the Kallman syndrome, HH with anosmia (unability to smell) results from deficient GnRH neuron migration to the hypothalamus during embryogenesis. Mutations in several genes including *ANO1 (KAL1)*, causing the Kallman syndrome, and for example *FGF8*, *SOX10*, *GNRH1*, *GNRHR*, and *KISS1R* have been linked with congenital HH (4, 65, 73). In addition to these HH cases, in rare cases mutations in the genes encoding for the pituitary hormones LH and FSH or their receptors have been identified as causes for absent puberty (77-79).

Compared to syndromes causing either severe delay or absence of puberty, the genetics of central precocious puberty (CPP) have been relatively less studied. The classification of precious puberty is somewhat arbitrary and has been traditionally defined as pubertal onset that occurs 2-2.5 SD before population mean (i.e. 8yrs in Caucasian girls and 9yrs in boys). Mutations in four genes, *KISS1*, *KISS1R*, makorin RING-finger protein 3 (*MKRN3)* and Delta-like 1 homolog (*DLK1)* have been associated with this phenotype (80). Intriguingly, gain-of-function mutations in *KISS1* and *KISS1R* have been linked with CPP, as opposed to the loss-of-function mutations in *KISS1R* causing absent puberty, consistent with the idea that kisspeptin release stimulates GnRH release and thus pubertal onset (39, 80-82). However, such mutations do not appear to be a common cause for CPP, likewise to *DLK1* mutations. Instead, several studies have highlighted the contribution of *MKNR3* to CPP (2, 80, 83). Contrasting the *KISS1* findings, the normal function of both *DLK1* and *MKRN3* appears to be suppression of pubertal onset, as deletions in these genes have been associated with CPP (84). Both genes also show maternal imprinting, meaning that only the paternally inherited copy of the gene is expressed (84). Still, the exact mechanisms whereby the genes contribute to CPP remain unclear.

2.2.3 Polygenic contribution to pubertal timing – insights from GWAS

Besides the insights gained from monogenic disease, technological, methodological and analytical advances in the field of genetics over the past decade have made it possible to conduct large-scale association studies aiming to identify genes that contribute to the individual variation in pubertal timing in the healthy population. The genome-wide association studies (GWAS) have provided a successful hypothesis-free method to probe the pubertal genetics in an unbiased manner, leading to the identification of hundreds of genetic loci affecting pubertal timing in both sexes. As opposed to the findings from the monogenic disease (rare variants exerting large effects on pubertal timing in a core set of genes that are essential for puberty and usually show little variation in healthy individuals, discussed in chapter 2.2.2), the GWAS have instead pinpointed genes that contribute to the regulation of pubertal timing in more subtle ways.

Importantly, the GWAS have potential to tell us about the effects of common sequence variation, present in most of us, affecting "normal" population variation in many traits. The type of sequence variation the GWAS generally utilize is single nucleotide polymorphisms (SNP) that are used as markers for genomic loci. Traditionally, the GWAS have utilized dense SNP-based arrays to associate common sequence variation (previously defined as the marker locus having minor allele frequency (MAF) >5%, though modern studies have often set the limit to >1%) with a trait (85, 86). The GWAS can be divided into two categories depending on the experimental setup. Under a case-control setting, the GWAS aim to detect differences in the frequency of genetic variants between two groups (e.g. those who have a disease and those who do not + other dichotomous traits). Secondly, the GWAS can be applied to study quantitative traits in large population cohorts: in this case they compare whether the carriers of a certain genetic variant differ in a statistically significant way from the rest (e.g. are they on average taller than others). For dichotomous traits the statistical method used to detect association between a SNP and disease is usually logistic regression, whereas for quantitative traits linear regression is generally used. The results from the GWAS have been generally proven robust and replicable once common, stringent standards for these experiments took place (85, 86). These include for example the genome-wide significance threshold (P<5e-8), that is based on the assumption that a human genome contains roughly a million independent genetic variants (87).

The era of GWASs began around 2005 when a study assessing age-related macular degeneration was published, reporting two single nucleotide polymorphisms (SNPs) that associated with the disease(88). It then took a few years before the first GWAS assessing pubertal timing were published in 2009 (89-92). These studies, consisting of 5000-25000 individuals associated sequence variation near *LIN28B* with changes in the timing of AAM, which together with a locus at chromosome 9 became the first loci where common genetic variation could be linked with pubertal timing (**Figure 3**) (89-92). Since then, the number genetic loci shown to affect pubertal timing has grown to ~400, with the latest paper consisting of genetic data from ~370,000 women (10). By constantly increasing the sample size, the GWASs have become more powerful to detect common variants with minuscule genetic effects. Despite the associations have generally proved to be robust, the effects of most individual loci are relatively small, and at the moment they have been estimated to

explain collectively ~7.4% of population variance and ~25% of the heritability of AAM(10). Consistent with the epidemiological data, many of the pubertal timing loci affect also body mass, reflecting the close relationship between pubertal timing and body size (93). Moreover, the pubertal timing loci often contain genes whose function seems to be conserved in evolution: for example genes like *LIN28B* and *VGLL3* have been shown to affect developmental timing and sexual maturation in species ranging from nematodes to fish, mice and humans (9,49). Though larger and larger GWASs will likely provide more hits in the future, the common variants that exert the biggest effects on pubertal timing have likely already been identified. Generally, the results from the GWAS emphasize pubertal timing as a complex trait. For example, although some of the hundreds of genes linked with the trait can be directly associated with GnRH-related pathways, most of the genes have been never linked with pubertal timing before the GWASs, stressing the important contribution these studies have made for understanding the genetics of pubertal timing.

Regarding the interpretation of the GWAS results, some limitations still exist. Although strong, replicable statistical associations with several genetic loci and pubertal timing now exist, in most cases we still do not know the true causal variant, nor the causal gene behind the association signal. In the case of pubertal timing it also must be kept in mind that most studies have been targeting AAM, and the results from these studies should not be generalized directly to be applicable in both sexes. Yet, smaller studies using Tanner staging, age at voice brake and the timing of the pubertal growth spurt as proxies of pubertal timing have made it possible to study also pubertal timing in men (16, 94, 95). By and large, the results from these studies overlap, and many of the pubertal timing associated genes show association in both sexes (13). For example, several different SNPs in the *LIN28B* locus have been robustly associated with pubertal timing in both males and females.

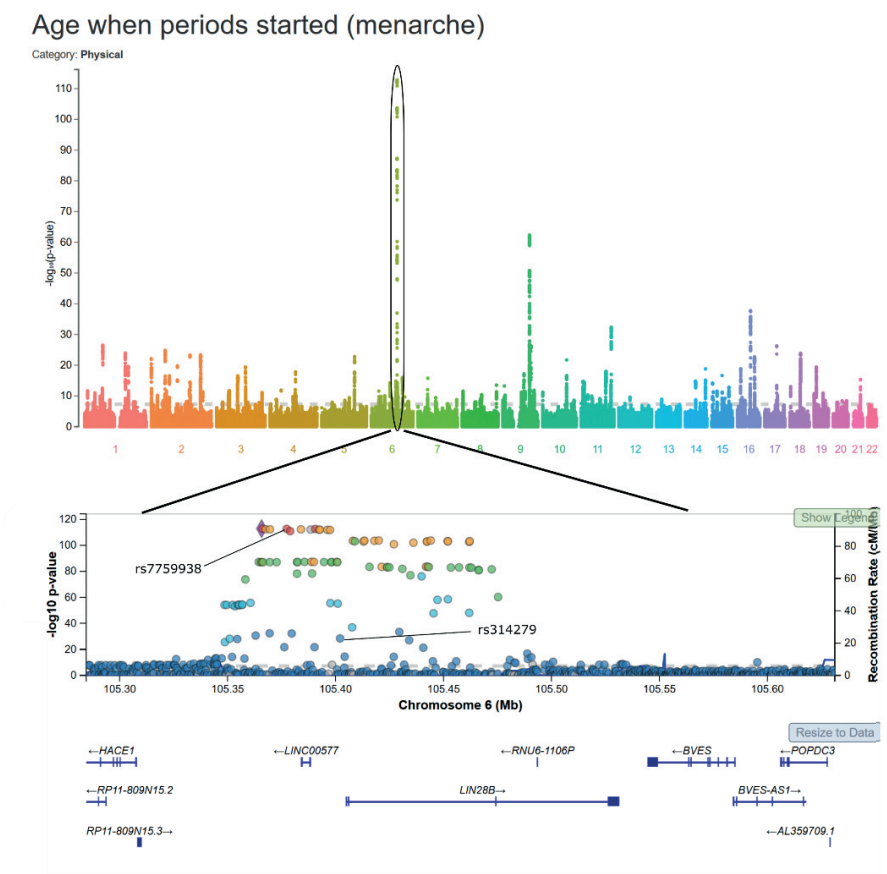


Figure 3. Visualization of the association of *LIN28B* locus with age at menarche (AAM). The upper panel shows a manhattan plot depicting chromosomal regions that show association with AAM in the UKBB cohort(19). The region close to *LIN28B* contains several sequence variants that show robust association with AAM (Black box). Notably the locus contains the most significant AAM associations out of all tested genomic regions, as indicated by the log10 p-value shown in the y axis. The lower panel shows a zoomed-in regional association plot of the *LIN28B* locus, showing the location of rs7759938, one of the markers associated with AAM, and rs314279, another marker studied in this thesis. The marker resides in a haplotype overlapping *LIN28B* anti-sense mRNA and *LIN28B* promoter region that seems to drive the association signal in the locus. Figure and data adapted from Oxford Brain Imaging Genetics (BIG) Server (v2.0, http://big.stats.ox.ac.uk/)

2.2.4. Epigenetic control of pubertal timing

Epigenetics refers to heritable mechanisms that affect gene expression without causing direct changes to the DNA sequence. These mechanisms are mostly related to changes in the DNA methylation and chromatin structure, caused by cellular mechanisms involving methylation enzymes, histone modifications and miRNA activity (96, 97). Heritable in this context refers to both meiotic and mitotic inheritance, meaning that the epigenetic changes can either be inherited from a parent to child, or from a parent cell to the daughter cells. In terms of puberty, this means two things. On one hand, some of the heritability in pubertal timing may be explained by meiotically inheritable epigenetic changes in the genome. On the other, epigenetic changes that are part of the normal ontogeny, occurring for example in the cells of the hypothalamus, may be crucial for the regulation of the pubertal onset.

Parental origin effects, referring to whether a sequence variant is inherited from the father or the mother, represent a subclass of heritable epigenetic mechanisms (98). This is generally referred to as genomic imprinting. It is an epigenetic phenomenon which means that only either paternally or maternally inherited copy of a gene is expressed, whereas the other allele remains silent following DNA methylation. Interestingly, several genes linked to monogenic forms of pubertal timing disorders have been shown to be imprinted, including *MKRN3* and *DLK1* (83, 84). Also GWAS have associated several imprinted loci with pubertal timing. A larger proportion of sequence variants associated with AAM seem to reside in imprinted genomic locations compared to all other studied complex traits (4.8% vs. 1.7%) (13, 65). The observation that many pubertal timing genes might be imprinted may have evolutionary background. It has been speculated that imprinting has emerged as a means to address the parental conflict of resource allocation. Maternally expressed genes should favor earlier puberty (to ease the parental load of mothers), whereas the normal function of paternally expressed genes should be to delay puberty (demanding more resources from the mother), like in the case of *MKRN3* and *DLK1* (83, 84). The gene that is the focus of this thesis, *LIN28B,* has not been shown to be imprinted in adult human tissues, but there is some evidence that it may be imprinted in the placenta, where the paternal allele is preferably expressed (99).

Besides some of the pubertal timing genes showing heritable imprinting, epigenetic changes at the time of pubertal onset in relevant tissues may be vital for understanding the molecular mechanisms that cause puberty. Pubertal timing involves a dramatic shift in the gene expression at the HPG axis, and such substantial changes in gene expression are thought to be driven largely by epigenetic mechanisms (100). In fact, epigenetic changes involving chromatin remodeling in the hypothalamus at the time of puberty have been shown to be extensive, causing either up- or downregulation of selected genes (37, 101, 102). For example, it has been shown that in rats changes in the methylation of the *Kiss1* promoter region, leading to activation of *Kiss1* expression, occur around the time of puberty (101). Despite these examples and the appealing concept that epigenetic regulation must contribute to pubertal onset, the research over this subject is still in its infancy and we are only beginning to understand the nature and causes behind the epigenetic mechanisms that affect puberty.

2.3 *LIN28B*, a special pubertal timing gene

2.3.1 Discovery of *LIN28B,* an evolutionarily conserved regulator of development timing

The studies that form the body of this thesis have been largely concentrated on understanding the function of *LIN28B*, a gene associated with pubertal timing via GWAS in humans. The next sections of the literature review will concentrate on describing the discovery, known function and some open questions related to *LIN28B* and its role in human development. Before the work for this thesis begun, relatively little was known about the normal function of *LIN28B.* In fact, the gene was identified in humans only few years before the GWAS discoveries. In 2006, a Japanese group reported cloning and characterization of a novel gene lin28-homolog B (*LIN28B*) that was overexpressed in hepatocellular carcinoma (103). However, already by the end of the decade, homologous genes *LIN28B* and *LIN28A* were under intensive study, after discoveries from various fields had linked the *LIN28* genes with cancer, growth and stem cell biology. Similarly to other genes in the *LIN28* family, *LIN28B* encodes for a RNA-binding protein containing a cold shock domain (CSD) and two retroviral-type (CCHC) zinc finger motifs (**Figure 4**) (103, 104). *LIN28* genes are found in all metazoans, and the essential domains of the proteins are highly conserved in evolution. No other proteins are known to contain such a combination of functional domains, suggesting the *LIN28* genes may have a special role in the biology of animals (103). The genes are highly expressed during embryogenesis in all studied animal species, ranging from nematodes and fruit flies to chicken and mice, and are also present in the embryonic stem cells from various organisms and in cell lines derived from somatic tumors (105-107).

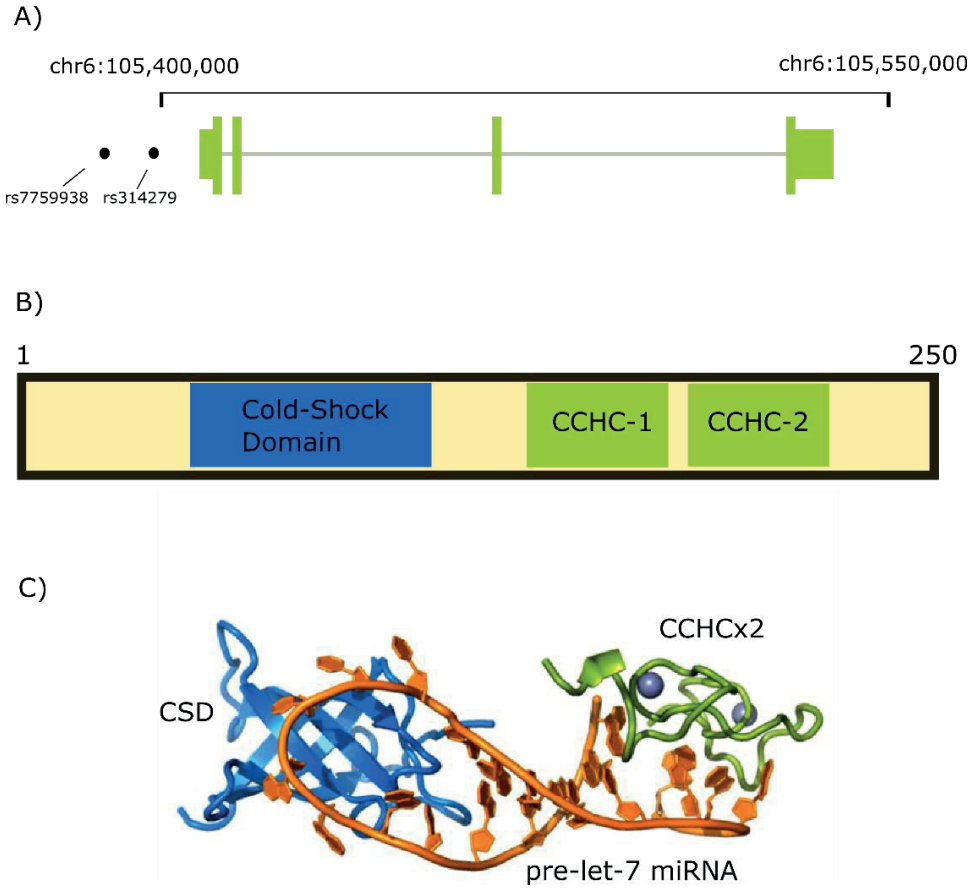


Figure 4. Illustration of the *LIN28B* gene and protein structures. A) Schematic depiction of the exon organization of *LIN28B* and locations of the SNPs associated with pubertal timing assessed in this thesis. Line above shows rough genomic position of the gene in humans, green boxes mark exons and dots show the relative location of pubertal timing associated SNPs. B) Illustration of locations of the major functional domains of the LIN28B protein (numbers refer to amino acids), and C) a picture of the interaction between pre-let-7 miRNA and LIN28-protein. CSD=Cold-Shock Domain. Image based on data from the UCSC genome browser (108) and an LIN28 image by Piotr Sliz, released under creative commons license.

Prior to the GWAS results, neither *LIN28B* nor the homologous gene *LIN28A* in humans had been thought to contribute to pubertal timing. Nevertheless, remarkably some studies had already linked the gene with developmental timing – though in a less complex organism than humans, namely the nematode *C. elegans* (11). In the 1980's, researchers in Boston

performed several forward genetic screens where they induced random mutations to the nematodes by chemical methods. When an interesting phenotype following the chemical treatment was observed, they mapped the genetic location of the mutation and isolated the likely gene behind the phenotype. One of these forward genetic screens targeted mutants that caused a "heterochronic" phenotype, meaning that the mutations affected the timing of developmental events. These screens led to the identification of several genes, including *lin-28*. The gene was named after the mutational *lineage* detected in the study - in this case *lineage 28* (11). The researchers observed that loss-of-function mutations in *lin-28* caused premature differentiation of vulval and hypodermal stem cells, and it was later shown that, conversely, a gain-of function mutation in *lin-28* caused an opposite phenotype, namely prolonged cell division in the same cells (11, 104). Importantly, the results from these early studies in *C. elegans* enforced the idea that *LIN28B* might truly be the causal gene behind the GWAS signals. Moreover, they suggested that the links between pubertal timing and *LIN28B* might be related to some fundamental, evolutionarily conserved molecular mechanisms that regulate developmental timing and ontogeny across the animal kingdom.

2.3.2 The cellular function of the *LIN28 genes: miRNAs, lin28/let-7* pathway and stem cell biology

The discovery of the heterochronic (=affecting the timing of developmental events) gene *lin-28* was soon coupled with a finding related to its cellular function, again based on work done in *C.elegans* (104, 109). In the nematode, another heterochronic gene, *lin-4*, had been identified, where a loss-of-function mutation caused partially opposite phenotypes compared to *lin-28* (104, 109). *lin-4* was shown to encode for a small, 21 nt RNA molecule that could regulate protein translation by binding to 3'UTR of targets with complementary sequences (110). The molecular cause behind the opposing phenotypes of *lin28* and *lin-4* mutants began to unravel when it was discovered that *lin-28* expression is temporally regulated by the *lin-4* RNA, acting through a *lin-4*-complementary element in the *lin-28* 3'UTR (104). Perhaps the most significant finding related to *LIN28* function however came in 2000 when another small RNA, *let-7* (lethal 7), was identified as a negative regulator of *LIN28* through a similar mechanism as *lin-4*, once more in *C.elegans* (111-113). When

overexpressed, *let-7* was shown to cause precocious differentiation of cells during larval stages, as opposed to the repeated cell division associated with *lin28* overexpression (111). As a notable exception, whereas *lin-4* was found only in the nematode, members of the *let-7* family were found in the genomes of a wide range of metazoans (114). Besides offering insights in the function of *lin28*, the discoveries of *lin-4* and the *let-7* were landmarks also in terms of the history of cellular biology: these showed the potential of small RNA molecules with conserved functions to affect gene expression, marking the discovery of microRNAs (miRNA) (110).

A few years after, a hallmark paper was published showing that human fibroblasts could be successfully transformed into induced pluripotent stem (iPS) cells (115). This paper, marking a completely new era in stem cell research, identified *LIN28* as one of the four factors that could be used in reprogramming human cells into iPS cells, a feature that has been lately shown to apply also to *LIN28B* (116). This again stimulated plenty of research addressing *LIN28* function in stem cells in more detail. Soon, *LIN28s* were shown to inhibit *let-7* miRNA biogenesis through direct binding to either pre-let-7 and/or pri-let-7 in mammalian cells (117-121). As *let-7s* were now known to repress *LIN28* translation, while *LIN28s* were understood to inhibit the biogenesis of *let-7* microRNAs, it became clear that such a two-directional switch could represent a major cellular mechanism governing stem cell renewal in multicellular organisms. Since then the concept of *lin28/let-7* as a cellular switch determining whether the cells continue to divide (situation when *lin28* dominates) or differentiate (*let-7s* dominant) has indeed been established as one of the key mechanisms that regulates the choice of stem-cell fate. In reality, the molecular features and the regulation of this switch in many cells is likely more complex than a simple "*LIN28* on – *let-7* off and vice versa" –model implies (122). Keeping this in mind, although a rough consensus of the molecular processes whereby the *LIN28*/*let-7* axis affects the regulation of cell division has been established (**Figure 5**), how this is related to more complex biological phenomena like pubertal timing remains poorly understood.

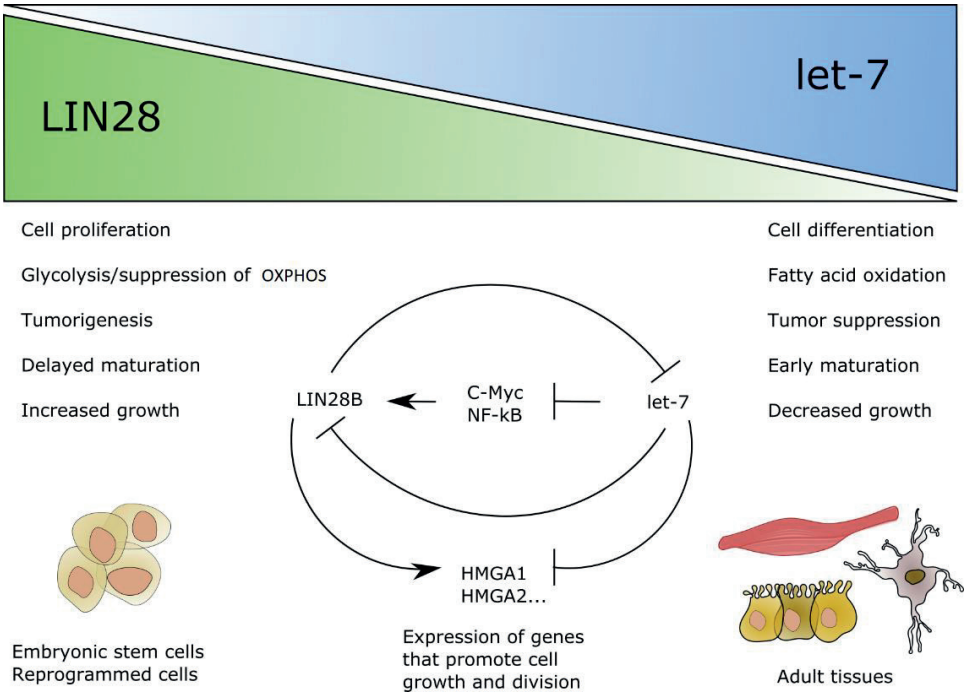


Figure 5. The main features of the *LIN28/let-7* axis and an overview of *LIN28B* regulation. Whilst *LIN28* genes are known to promote cell growth and division, *let-7* miRNAs are mainly expressed in differentiated cells. At the moment it is thought that the axis represents a major switch that determines which gene expression patterns dominate in cells. Generally, the situation where *LIN28* expression dominates is considered to promote cell proliferation and glycolysis at the expense of mitochondrial oxidative phosphorylation (OXPHOS). *C-Myc* has been established as a major driver of *LIN28B* expression in cancer cells, and *NF-kB* has been shown to drive *LIN28B* expression upon inflammation. Many downstream effectors of *LIN28B* including HMGA1 and HMGA2 drive conformational changes in chromatin organization affecting the expression of several genes. Figure based on reviews by Thornton&Gregory 2012 and Balzeau et al. 2017 (123, 124)

2.3.3 *LIN28B* as a cancer gene

Cancer is fundamentally a disease of uncontrolled cell division. *LIN28B* was discovered as a gene that was overexpressed in hepatocellular carcinoma, and *LIN28B* overexpression has been since observed in many different types of tumors (103, 107, 123). As it was also known that *let-7* miRNA expression is usually downregulated in cancer cells, the *LIN28/let-7* axis has subsequently been studied in pretty much all possible cancer subfields. Cancer research represents a significant segment of the whole field of biomedical study, and many research

projects have been focused on addressing the genomic mechanisms related to cancer incidence and progression. Therefore, it is not surprising that besides the fields of developmental and stem cell biology have contributed a lot to the knowledge about the molecular mechanisms of *LIN28* function, a plethora of research concerning *LIN28*/*let-7* axis has been made from the angle of how this axis contributes to cancer formation. Generally, these studies have reported upregulation of *LIN28*/downregulation of *let-7* miRNA in cancer cells (107, 123). Since *LIN28/let-7* axis contribution to cancer has been widely studied, several excellent reviews are available around this subject (107, 123, 124). The variety of cancers where *LIN28/let-7* axis dysregulation has been reported is extensive. Examples of these include breast cancer, cervical cancer, gastric cancer, hepatocellular cancer, neuroblastomas and prostate cancer (103, 123, 125-128). At the moment, multiple lines of evidence suggest that expression of *LIN28* genes may contribute for example to the formation of cancer stem cells (CSCs), metastasis, and resistance to chemotherapy and radiotherapy (123).

2.3.4 Insights to *LIN28* function from animal models

Despite the studies on *lin28* in C. elegans providing support for the idea that *LIN28B* might indeed affect the development timing in humans, very little direct evidence linking *LIN28B* with puberty existed at the time of the principal GWAS findings. In fact, as is the case for most GWAS loci, although *LIN28B* was the nearest gene close to the associated SNPs, it remained somewhat uncertain whether the actions of this gene were truly behind the observed association signal. However, soon after the original GWAS findings,a landmark study, examining the function of *Lin28A* in mice added evidence for *LIN28s* capability to affect developmental timing (129). This study revealed how "leaky" overexpression of *Lin28A* yielded similar phenotypes in mice as linked with *LIN28B* in humans: mice engineered to overexpress *Lin28A* showed delayed maturation, as well as increased size, reflecting the human phenotypes associated with *LIN28B* (129). The leaky overexpression was based on introducing an additional copy of *Lin28A* into a *Collagen 1a1* locus. The transgenic mice showed proportional overgrowth of all organs, suggesting that an endocrine- or metabolism-related mechanism might be the cause for the increased size in these mice. Finally, the authors showed that *Lin28A* overexpression led to an enhancement in glycolytic metabolism

both in the mice and a myoblast model, suggesting this as the potential reason behind the increased growth related to the overexpression of the gene (129). A year later, the same group of researchers reported that the *Lin28/let-7* axis can modulate glucose metabolism in mice, by regulating multiple genes belonging to the insulin-PI3K-mTOR pathway (130). Reportedly, both *Lin28A* and *Lin28B* transgenic mice manifested increased glucose tolerance, and were resistant to obesity, whereas muscle-specific *Lin28a* knockout mice showed opposing phenotypes. However, whether *LIN28B* could contribute to metabolic traits in humans remained still unclear, representing one of the questions addressed in this thesis.

Further research utilizing the *Lin28A/Lin28B/let-7* mouse models has since then led to some other key discoveries. Based on conditional knockout studies in mice, embryonic *LIN28B* expression seems to have the greatest potential to regulate growth, since *LIN28B* knockout after the fetal stages did not affect mouse size significantly (131). A separate study linked *Lin28A* with tissue regeneration: mice engineered to overexpress *Lin28A* displayed evidence of improved tissue repair (132). Captivatingly, the mice showed increased hair growth after shaving, and improvements in digit and ear pinnacle repair following injury. Up to date, similar tissue repair- phenotypes have not been reported for *LIN28B*. Most recently, sex-specific growth patterns following *Lin28B* knockout in mice have been reported: it appears that *LIN28B* affects growth primarily in male mice, reducing both body length and weight (131, 133).

In addition to the mice studies, LIN28/let-7 system has been explored in rats. The rat studies have been based on models that have examined *lin28b* expression in the hypothalamus under normal and experimental physiological conditions. These studies have reported that *lin28b* levels decline towards sexual maturation in rat hypothalami, suggesting that this might somehow be permissive for pubertal onset (134). Nevertheless, also these studies have offered relatively little mechanistic insight into the molecular function of *Lin28b* at the time of puberty.

On top of these observations, some efforts have been made to study *LIN28* also in other animals, yielding partly similar, but also somewhat contrasting results compared to the

rodent models. For example, early reports from both the frog *Xenopus laevis* and zebrafish *Danio rerio*, claimed that knockdown of *LIN28* by morpholinos might cause embryonic lethality (135, 136). Notably, this type of knockdown method, based on injecting antisense nucleotides to embryos differs considerably from the permanent and conditional mouse knockout models, as will be discussed in the section 2.4.7. Contrastingly, for example studies in *Drosophila melanogaster* have indicated that *lin28* mutants are viable, though have defects in oogenesis and muscle formation (137).

Overall, the various model organism studies have significantly advanced our understanding of the function and potential pleiotropy of *LIN28* genes in the development of vertebrates. Yet, most of these studies have not been specifically designed to address the molecular mechanisms by which *LIN28B* genes control sexual maturation. Here, information remains still sparse, representing a central gap in knowledge that this thesis aims to fill.

2.3.5 *LIN28B,* a pleiotropic gene affecting several phenotypes in humans

After the association of *LIN28B* with pubertal timing, it became soon apparent that the gene has some distinctive characteristics that make it an interesting subject for further research. For instance, the magnitude of the effects associated with genetic variants in the *LIN28B* locus is relatively large, especially in GWAS terms: on average one pubertal timing-advancing allele (T) for SNP rs7759938 causes a ~1.5 months advancement in AAM (10, 12, 13). Such a large effect on pubertal timing is unusual for common variants. Both pubertal timing advancing and pubertal timing delaying alleles in the locus are present in substantial frequencies in all studied modern populations: the frequency of the T allele at rs7759938 is ~ 67% in the Finnish population, causing ~45% of Finns to be homozygous for it. Respectively, ~44% are heterozygous carriers of the T and pubertal timing delaying (C) allele, and ~11% are homozygotes for the C allele (138).

A GWAS addressing the genetics of pubertal growth spurt, performed by our research group later showed that the effect on pubertal timing applies to both males and females (9). Since then it has become evident that the effects *LIN28B* exerts on human growth are widespread,

and not limited to pubertal timing. Besides puberty, sequence variation in the *LIN28B* locus has been currently linked with several other traits. These include well studied physical characteristics such as height, weight, waist and hip circumference, more specific traits such as 2D4D finger length ratio, and bone mineral density (BMD), but also disease-related phenotypes like intraocular blood pressure, neuroblastoma, insomnia and depression that have been linked with *LIN28B* only recently (15-18, 139-146). The associations are summarized in **Table 2**. The directions of the effects emphasize the complexity of the effects *LIN28B* has on human development. The same variants that delay pubertal onset (and based on mouse models show an effect that is likely related to increased *LIN28B* expression) cause an increase in human height, weight and waist and hip circumference. At the same time, they also cause an increase in 2D4D finger length ratio, which has traditionally been thought as a marker of prenatal testosterone exposure (high 2D4D = longer index finger compared to the ring finger, a feminized phenotype) but also decrease bone mineral density. Disease-wise, the variants that associate with increased *LIN28B* expression seem to decrease the risk of depression and insomnia, but they also increase intraocular pressure (the main risk factor for glaucoma), and the risk for schizophrenia and neuroblastoma.

The work done in mice suggests that *LIN28B* contribution to cellular metabolism might be the cause behind some of the phenotypes associated with *LIN28B* also in humans (130). Notably, most of the traits associated with *LIN28B* in GWAS are also a subject to hormonal regulation (e.g. height, pubertal onset, finger length ratio etc.). Furthermore, there is some evidence of sex-specific effects of *Lin28B* in mice as well as in humans, and neonatal estrogenization has been shown to affect *Lin28B* expression in the rat hypothalamus (133, 134). Therefore, the possibility that *LIN28B* might be involved in hormone-dependent regulation of vertebrate development in more central ways than perhaps previously appreciated should not be excluded, but rather explored in more detail.

2.3.6 The prevalence of protein coding mutations in *LIN28B* in humans

The current evidence suggests that *LIN28B* contribution to pubertal timing in the healthy population is related to gene expression levels, and does not stem from protein coding mutations. For example the data from the Exome and Genome Aggregation databases (ExAC

and GnomAD), indicates that truncating mutations in *LIN28B* are very rare (138, 147). In the ExAC database, only ~1:60,000 individuals seem to carry a protein truncating variant in *LIN28B*, while based on the expected mutation rate, the database might contain at least 8 such variants (147). Likewise, neither database reports any individuals that are homozygous for *LIN28B* truncating mutations. In humans, such depletion of severe damaging variants is often typical for genes affecting neurodevelopmental disorders (148). There are also no known high-frequency missense variants in the gene that co-segregate with the pubertal timing associated SNPs, and thus such variants in *LIN28B* gene are unlikely to explain the GWAS results.

Table 2. List of traits associated with *LIN28B* in humans in GWAS. The table shows the most significant SNP, the effect size and details of the study. For clarity, the direction of the genetic effect is listed separately always for the allele that increases *LIN28B* expression.

Trait	SNP(effect allele)	Beta of OR	P-value	Sex	Population	N	Direction of effect	Study
Age at menarche	rs395962 (T)	0.127	2x10-213	F	European	329,340	Delay	Day et al. 2017
Age at voice drop	rs9391253 (T)	0.058	1x10-23	M	European	55,871	Delay	Pickrell JK et al. 2016
Late pubertal growth (14 to adult)	rs7759938 (T)	-0.094	4x10-9	M+F	European	8,863	Increase	Cousminer D et al. 2013
Menopause	rs395962 (T)	0.079	5x10-16	F	Japanese	43,861	Delay	Horikoshi M et al. 2018
2D4D digit ratio	rs314277 (A)	0.48	1x10-21	M+F	European	12,636	Increase	Warrington NM et al. 2018
Height	rs7759938 (T)	-0.045	8x20-31	M+F	European	183,653	Increase	Lango Allen H et al. 2010
Waist circumference adj. for BMI	rs7759938 (C)	0.030	3x10-15	M+F	Multiethnic	200,452	Increase	Graff M et al. 2017
Hip circumference	rs7759938 (T)	-0.028	2x10-13	M+F	Multiethnic	223,733	Increase	Shungin D et al. 2015
Depression	rs1475120 (G)	-0.029	4x10-9	M+F	European	478,240	Decrease	Hyde CL et al. 2016
Feeling guilty	rs12528131 (A)	6.42(OR)	1x10-10	M+F	European	373,380	Decrease	Nagel M et al. 2018
Heel bone mineral density	rs1475120 (G)	-0.013	3x10-13	M+F	British	426,824	Decrease	Morris JA et al. 2018
Intraocular pressure	rs7759938 (C)	0.083	5x10-9	M+F	European	115,486	Increase	Gao XR et al. 2018
Neuroblastoma	rs17065417 (A)	1.43(OR)	5x10-9	M+F	Multiethnic	11,862	Increase	McDaniel LD et al. 2017
Schizophrenia	rs160593 (G)	1.06(OR)	8x10-9	M+F	Multiethnic	26,026	Increase	Li Z et al. 2017

Similarly, it appears that protein coding mutations in *LIN28B* are not a common cause for rare pubertal timing disorders. For example, a study including 145 Finnish patients diagnosed with delayed pubertal timing, found no protein coding mutations in *LIN28B* (149).

2.3.7 Summary of the findings – what we currently know about *LIN28B*

To summarize, several studies have connected *LIN28B* with pleiotropic effects on metazoan ontogeny and adult phenotypes. The gene has been linked with the regulation of several aspects of growth, e.g. organismal size and timing of sexual maturation, and it is known to contribute also to several diseases ranging from cancer to mental disorders in humans. Currently, knowledge about the biological mechanisms and pathways related to *LIN28B* is based on integration of results from various scientific fields. These fields include developmental biology, stem-cell biology, cancer research, various animal models and large scale human studies. Based on these studies, it is now thought that the major function of *LIN28B* is to control cell division and metabolism through the *LIN28/let-7* axis. Yet, how and whether this molecular function affects the variety of phenotypes associated with *LIN28B,* remains largely unsolved.

2.4. From GWAS to biology

2.4.1 The challenge: the gap between the associations and biological knowledge

The era of GWAS has marked a substantial advancement for understanding the genetics of complex traits. We know now that variation in thousands of genetic loci contributes to the heritability of complex traits (20, 86, 150). Even though the GWAS have provided statistically significant and replicable associations with thousands of genetic loci and phenotypes, the interpretation of these findings is however currently limited. At the moment, a big challenge relates to moving from the GWAS results into understanding the biology behind these associations. This challenge essentially provided the motivation also for this thesis.

In order to make the leap from the genetic associations into explaining the biology, it is necessary to understand some basic principles of the GWAS design. Although the GWAS nowadays include from hundreds of thousands to millions of directly genotyped variants, these still represent only a fraction of all variants that exist in the human genomes (20, 85, 86). With the help of this fraction it is still possible to gain information about most of the genome: genetic variants are generally not inherited independently, but form haplotypes e.g. sets of genetic variants that are inherited together. In essence the markers used in the GWAS, SNPs, can pinpoint a haplotype (a piece of chromosome, a genetic locus) that associates with a trait, instead of directly pinpointing a causal sequence variant, or a gene. Therefore, in order to identify the causal variant, the biological mechanism and the causal tissue behind the association signals, further studies are required (20, 85, 86). Efforts including statistical fine mapping of the genetic locus (i.e. characterizing all sequence variation in a given locus by sequencing and/or imputation, followed by statistical analysis that leads to prioritization of variants) or prioritization of variants based on functional annotations of the human genome can be used to tackle the challenge of identifying the true causal variant (151, 152). This in turn has potential to lead also to the identification of the causal gene in the locus, which is similarly challenging. Traditionally, it has been the convention that the genes nearest the associated SNP have become the symbol for these

associations, even when robust evidence supporting their causality has been lacking. In the case of *LIN28B* and pubertal timing, this assumption has proven likely correct, but for example in the case of the most prominent obesity locus, a lot of research was first concentrated around FTO as opposed to the IRX3/IRX5 genes that are currently considered causal for the obesity phenotypes (153, 154). In essence, without careful analysis and further studies it often remains impossible to know from what type of variant, gene and biological effect the association stems from.

2.4.2 Essential features of genetic loci contributing to complex traits like pubertal timing

Famous studies by RA Fisher showed how in contrast to Mendelian or monogenic traits, inheritance of complex traits such as height and pubertal timing could be rather explained by the interaction of many genes and environmental effects (155, 156). Similarly, the theoretical model shown in **Figure 6**, as well as the results from several GWAS imply that most genetic variants affecting complex traits are fundamentally different compared to the variants that have been linked with monogenic disease (86, 155). Whereas the variants associated with monogenic disease are usually rare, having large effects on certain traits, the variants associated with complex disease are generally common having relatively small effects (85, 86, 150, 155). This difference largely stems from the biological properties of the associated variants. Mendelian disease is often caused by severe protein truncating variants or genomic lesions including insertions, deletions and translocations of genetic material that can have dramatic fitness consequences. In contrast, most of the GWAS associated loci contributing to complex traits and to normal population variation reside in the non-coding parts of the genome. In theory some of these markers might simply tag the effects of sequence variants affecting protein structure, but for most loci this seems not to be the case (86). Instead, these complex disease loci most often affect gene regulation (86). This has practical consequences also in terms of studying the GWAS loci: it should be carefully considered whether it is potential to gain insight into the GWAS associations for example from an animal model that utilizes protein truncation to study the function of the gene, or whether this type of a model may result in a phenotype that is totally unrelated to the GWAS association.

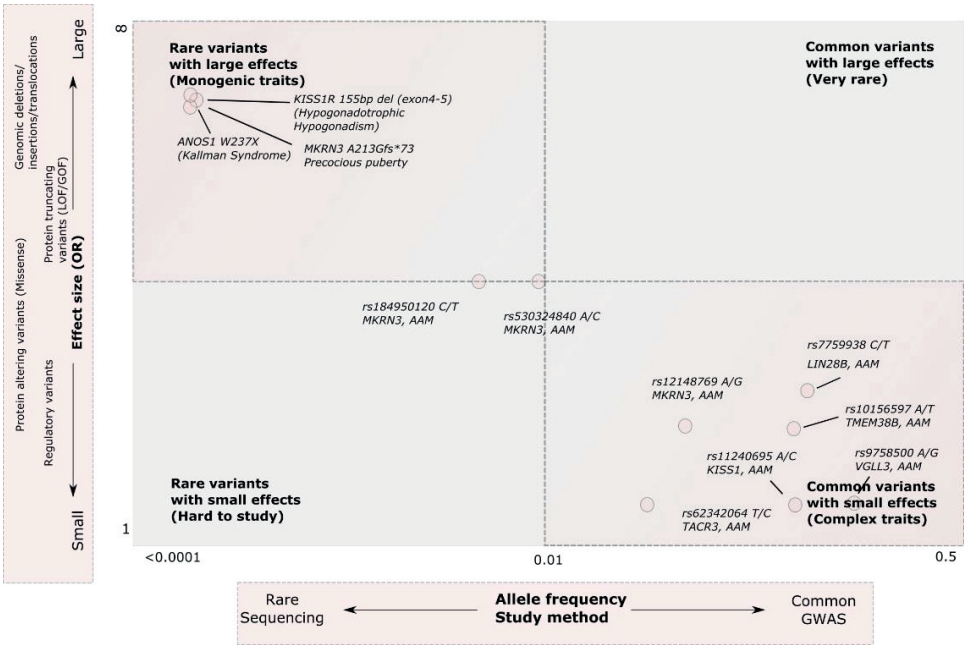


Figure 6. Illustration of the fundamental differences between Mendelian and complex traits. Whereas the syndromes that cause absent of precocious puberty are often caused by rare mutations that have large fitness consequences, the normal variation in pubertal timing between healthy individuals is thought to be largely due to aggregation of small genetic effects, often regulatory in nature, around the genome. Interestingly, pubertal timing gene *MKRN3* gene shows a full spectrum of variants in the human population, ranging from common variants with small effects and rare variants with intermediate effects to unique variants causing precocious puberty. Potentially, studies that rely on sequencing data have potential to highlight an increasing number of rare variants that may have small and intermediate effects on complex traits. Figure based on illustrations by McCarthy et al. 2008, Manolio et al. 2009 and Bush 2012 (85, 157, 158).

Another difference between monogenic and complex traits is that instead of clustering around a well-defined set of genes, the association signals in complex traits are spread across the genome (159). For many complex traits, variants having the largest effects may be modestly enriched in specific genes or pathways, but the vast majority of the association signals reside close to genes without an obvious connection to disease or trait, like in the case of *LIN28B* and pubertal timing (159). The effects of individual GWAS loci are generally small, and when considered collectively they have usually explained only a fraction of the genetic variance for a given trait. In the beginning of the GWAS era this was referred to as

mystery of missing heritability (157). This mystery has since been partly resolved through analyses which have shown that common SNPs with effect sizes below the genome-wide significant threshold account for a large proportion of the heritability of many traits (160, 161). This is reflected in the statistical phenomenon related to GWAS: as the sample size increases, more statistically significant findings have been found, but at the same time the effect size of the newly discovered loci has decreased. Collectively, these findings imply that there indeed exists a huge number of genetic variants that have small effects on complex traits. This has led to a hypothesis called an "omnigenic" model (159). The model states that a lot of the heritability for a given trait might be explained by the actions of genes that are not directly related to some pivotal pathways, due to interconnected gene regulatory networks. Theoretically, within a tissue, any change in gene expression might have consequences for the expression of other genes, although a gene may not always "directly" affect pathways that are directly responsible of formation of a given trait (159).

On the other hand, rare genetic variants with larger effect sizes may also contribute to genetic variance in complex traits. This has shown to be true especially for severe neurodevelopmental diseases such as autism and schizophrenia (162-166). Since on average every human carries 60-70 novel sequence variants of which 1-2 are protein coding (147), it is likely that in for every human generation also new variants that contribute to complex traits emerge. Since human population has increased exponentially over the past century, also the number of rare and private variants has increased. This is reflected by the observation that huge number of rare and private variants now exist in public databases that are based on sequencing thousands of humans (138, 167). Theoretically, the effects of these variants range from minuscule to Mendelian effects for an individual, and inevitably such variants are likely to explain a proportion of the heritability of complex traits. Therefore, whole exome and whole genome sequencing studies (WES and WGS) have potential to increase our understanding of the contribution of rare variants to complex traits. For example the results from the UK10K consortium and from the ongoing FinnGen project, including sequencing data from thousands of humans, have led and will lead to identification of several rare and low frequency variants that associate with complex traits, whilst these would have been impossible to detect by methods based on genotyping arrays (20). The complication is that even with the help of huge cohorts and in the presence of a true

biological effect and family information, it is statistically challenging to associate a rare variant with small effects with a trait in a robust manner (**Figure 6**). Despite this complication, recently, analyses of whole genome sequencing data have shown that such rare variation, that has previously been left understudied, may account for a large proportion (~50%) of trait heritability, essentially solving the problem of missing heritability of complex traits (168).

2.4.3 eQTL analysis and publicly available genetic databases may help to interpret GWAS results

Functional genetic variants can be roughly divided into two categories: 1) Those that cause a functional change in protein structure and 2) those that affect the expression level of a gene either globally or in a specific tissue. As discussed in the previous chapters, it is now thought that since most GWAS hits reside outside coding regions, they most likely tag variants affecting gene expression. To assist interpreting the biology behind each GWAS locus, characterizing the regulatory elements that reside in the intergenic and intronic locations in the human genome is necessary (169). Even without exact knowledge about the location of such regulatory elements, examining how genetic variation affects gene expression may provide vital hints for understanding the biology behind the associations. Usually, this is studied through expression quantitative trait locus (eQTL) analysis. eQTL refers to a genetic locus that affects the mRNA expression levels of a gene or several genes in within a tissue. In the past, eQTLs were identified in microarray studies assessing transcript abundance, but these have been largely replaced with RNA-seq based methods. Generally, SNPs that have been associated with a trait through GWAS have been shown to be enriched for eQTLs, consistent with the hypothesis that many GWAS loci affect gene regulation (170).

Cornerstones for modern genetic research are publicly available genetic databases. These range from sequence browsers and variant databases to databases containing information about gene expression and published genetic associations. A central gene expression database for this thesis was made available by the Genotype-Tissue Expression (GTEx) project (169). The GTEx database has greatly facilitated the study of human gene expression, by providing the scientific community with gene expression data from multiple different

human tissues. Importantly, the database contains also genotype information, making it possible to study whether, and in which way the variants identified through GWAS affect gene expression. At the moment, the database consists from hundreds of individuals and thousands of samples from a diverse set of human tissues (169).

Besides gene expression, understanding the phenome-wide consequences of a variant may greatly elucidate the biology of a GWAS finding. Phenome-wide association study (PheWAS) databases (essentially based on multiple GWAS on different traits in the same cohort(s)) contain association results for a single genetic variant against hundreds, if not thousands of traits (19, 171). Various databases that allow searching for phenome-wide effects for a particular SNP that are emerging (for example those based on the UK biobank- and the FinnGen projects) have thus vast potential to offer new insight into studying the GWAS loci.

2.4.4 Model organisms in studying the GWAS loci

Together with the data that is available from the genetic databases, various model organisms represent another fundamental asset for genetic and biomedical research. The common evolutionary background for all living organisms makes it possible to examine the function of genes in one species, and to infer how they function in another. Although the metazoans show extensive phenotypic variation, this variation largely stems from fundamental molecular mechanisms that are shared with all living creatures. The model organisms used to study the function of human genes range from bacteria and yeast to primates. In the past, many advances in the field of biosciences have been gained through the study of model organisms such as flies and mice.

In its most basic form, the genetic studies of model organisms have involved studying gene function by mutagenesis or studying the spatiotemporal expression patterns of a gene. Traditionally, the model organism studies have involved knocking out or overexpressing a gene and studying its effects on phenotypes. Technological advances have made it possible to refine these methods so that nowadays it is for example possible to prepare models that introduce a specific nucleotide change, or affect gene expression in selected cells. Different animal models have proven indispensable for studying the effects of the genes associated

with monogenic disease, often providing vital understanding about potential disease mechanisms. Although the GWAS loci identify variants that are partly distinct in nature compared to the monogenic disease loci, in many cases various animal models have still provided vital insights into the GWAS results. For example in the case of *LIN28B*, studies performed in nematodes and mice have been extremely informative (11, 129, 131).

2.4.5 Zebrafish as a model organism

To study the molecular mechanisms linking *LIN28B* with growth and sexual maturation, we have utilized zebrafish as a model organism in this thesis. Zebrafish are small, tropical freshwater fish that have become an increasingly popular model in genetic research over the past decades. The "founder" of zebrafish research, George Streisinger originally aimed to find a vertebrate model that would be simple to maintain under laboratory conditions, and identified zebrafish to suit his needs (172). Several aspects of zebrafish make them suitable for genetic studies. They are easy to breed and generate a large number of offspring, the maintenance of the fish is relatively straightforward and cheap, they develop quickly, and the external fertilization and transparent embryos make them amenable for genetic manipulation (172). An overview of zebrafish life cycle is presented in **Figure 7**. These central benefits of the fish have established zebrafish a one of the key models that are used in genetic research at the moment. The number of studies per year using zebrafish has risen from ~1000 in year 2000 into ~5000 by 2016, reflecting this popularity (173). The zebrafish have been extensively used as model for vertebrate development, tissue regeneration, behavior and several diseases, being also a popular model in toxicological studies. Several different tools for genetic manipulation of zebrafish have been developed (174). Overall, data from mutant screens, reverse genetics and transgenic zebrafish models has led to substantial advances in understanding the vertebrate physiology and development.

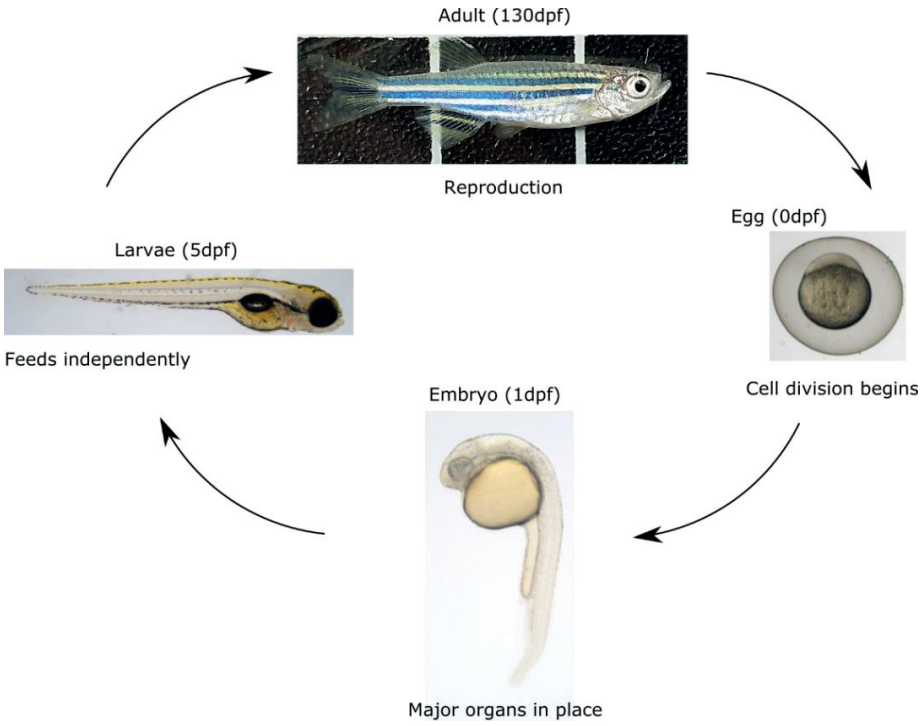


Figure 7. Zebrafish life cycle. The fish develop rapidly, and embryo transparency together with external fertilization make zebrafish amenable for many types of genetic analyses.

Many fundamental principles of growth have been conserved in the evolution of vertebrates, and for example the molecular pathways underlying development of the hypothalamus have been conserved in between fish and mammals(29). Also the function and basic structures at the HPG axis are built on similar principles as in mammals(29). The genome of the zebrafish has been sequenced, showing easily identifiable orthologues for most mammalian genes (175). Some notable genetic differences between the zebrafish and mammals however do exist. Teleosts (bony fish, including 96% of extant fish) such as zebrafish have undergone a genome duplication event, resulting in additional copies of many genes compared to mammals (175). For instance, the latest release of zebrafish genome contains three annotated *lin28*-genes: two orthologues of mammalian *LIN28A* in chr16 and chr19, and one orthologue for *LIN28B* in chr20 (108, 136, 175). The other genetic feature where the zebrafish differ significantly from mammals is the lack of distinct sex chromosomes (175, 176). Some genetic evidence supports the existence of autosomal loci that contribute to sex

determination, but also factors like water temperature and fish density have been suggested to affect sex determination of fish (176). These genetic features may underlie some unusual abilities that the zebrafish have compared to mammals, like the extensive regenerative ability to repair lesions after tissue damage (177). Despite these features of zebrafish, the basic molecular mechanisms governing processes like cell proliferation and biological principles controlling growth have been widely conserved between zebrafish and humans. Although compared to humans zebrafish appear to tolerate damaging mutations in key genes affecting reproductive capability, the conservation of the role of the HPG axis in the regulation of sexual maturation and reproduction in zebrafish was one of the major reasons why it was chosen as the model organism for the studies constituting this thesis (29, 178, 179).

2.4.6 Ethical aspects regarding zebrafish use in research

The ultimate goal of biomedical research is to understand how living systems work. To this end, the information gained from various animal models has proven crucial. The legislation concerning animal research is currently strict, and ethical aspects regarding all animal models, as well as animal welfare should be carefully considered before any experiment. Although in reality the experimental choices always include balancing between scientific gains and minimizing animal use, several recommended principles regarding the design of animal work exist, like the 3Rs: Replacement, Reduction and Refinement (180, 181). The use of zebrafish in research instead of mice and rats is an example of the Replacement principle. Zebrafish is considered as a simpler vertebrate than mammals, motivating the use of fish as a replacement for mice whenever possible. A good example of such a situation is research tackling the fundamental molecular pathways related to vertebrate ontogeny, like the studies forming the base of this thesis. The zebrafish have also some other advantages compared to many other animals. The oviparity of the fish (meaning that they lay eggs and hence embryonic development occurs outside the mother) is another feature of zebrafish that makes them especially well-suited for research that is focused on embryonic development: it is easy to get zebrafish embryos without harming the mother. This advantage together with the transparency of the embryos simplifies studying the immature forms of the fish, as a substitute for having to rely on adult individuals, which is another

preferred principle in animal research. The experiments done in fish under 5dpf (roughly the age when the zebrafish start to rely on external food sources instead of the nutrition provided by the yolk sac) are not subject to legal regulation and do not require an animal experiment permit in the EU (Directive 2010/63/EU). The use of fish older than 5dpf requires an animal experiment license, and to protect fish welfare and safety, the EU laws concerning animal experiments strictly regulate the housing conditions, anesthetic procedures and termination of zebrafish (EU document 32014R0377).

2.4.7 Genetic manipulation of zebrafish

Zebrafish have proven very amenable to several types of genetic analysis because of their special features discussed in chapter 2.4.5. In the past, zebrafish have been widely utilized in forward genetic screens in which random mutations were induced to fish, and upon detection of an interesting phenotype the causal genetic alteration was then mapped leading to the identification of several novel genes and pathways affecting vertebrate development (182). The classical methods to induce mutations in these forward genetic screens have involved radiation, chemical mutagens such as N-ethyl-N-nitrosurea (ENU), retroviruses and transposons (174, 182).

Besides forward genetic screens, particularly the easy access to zebrafish embryos has made reverse genetic screens in the organism increasingly popular once suitable tools became available. In contrast to the forward screens, the reverse genetic approaches aim at modifying a known genomic target instead of inducing random mutations, and subsequently following the phenotypic consequences. Examples of such methods are morpholino oligonucleotides (MO) and synthetic mRNA injections, Target Induced Local Lesion in Genomes (TILLING), zinc finger nucleases (ZFNs) and the increasingly popular CRISPR-Cas9 approach (CRISPR stands for Clustered Regularly Interspaced Short Palindromic Repeats) (174, 183). In this thesis, the studies involving zebrafish have been based on three different reverse genetic approaches, including gene knockdown by morpholino oligonucleotides (MO), gene overexpression by synthetic mRNA injections, and gene knockout by CRISPR-Cas9 technology.

MO are synthetic oligonucleotides that work in anti-sense fashion by binding to target mRNAs and subsequently inhibiting mRNA processing or translation (184). MOs, leading to transient gene silencing have been extensively used in studies targeting early development since they are easy to apply by microinjections into fertilized zebrafish eggs. However, like all anti-sense technologies, in many cases also MOs have proven to cause unwanted off-target effects, and only a fraction of the phenotypes associated with MO use have been replicated in studies using different mutational strategies (184, 185). Accordingly, the use of MOs requires great care and with the invention of novel mutagenic strategies has become less popular. The complementary method to MO induced gene silencing, synthetic mRNA injections, aimed to overexpress a gene seems to be better tolerated by the cells compared to MOs, though like in the case of MOs these may also cause unwanted cell toxicity if injected in two high concentrations (186).

CRISPR-Cas9 induced mutagenesis has meant a revolution for the genetic manipulation of organisms (187). The method is based on a RNA-directed endonuclease complex that generates double stranded breaks (DSB) in to DNA, representing a prokaryotic immune defence mechanism against viruses (187). The endonuclease consists of two main components, RNA sequence matching a genomic target site and Cas9 protein that is capable of cutting DNA. The genetic manipulation methods in zebrafish using this system usually involve generating a guide RNA (gRNA) that targets a specific genomic sequence, and co-injecting this with RNA molecules encoding for Cas9 protein into 1-cell stage embryos (183). In its basic form, the CRISPR-Cas9 method generally results in creation of a DSB that is quickly but inaccurately repaired by the cells. Usually this results in small deletions/insertions/mismatches in the target site, disrupting the target gene. However, the method can be used also to induce specific mutations into the target site (gene knock-in) (183). Despite the reported high specificity of the method in creating the desired genetic lesion, similarly to other mutational methods, potential off-target effects of the CRISPR-Cas9 technique have to be considered when designing an experiment (174, 183, 187). In practice, a couple of simple measures to control for CRISPR-Cas9 off-target effects exist. First, one may create multiple mutations to the same gene, using different target sequences, and check whether these have similar phenotypic consequences. Secondly, in case of using a model organism, the resulting genetic lesion should not be studied in the F0 generation, but

F0 organisms should be rather outbred to wild-type background before assessing the phenotype. Despite these caveats, due to its reliability, low cost, relative ease and high specificity CRISPR-Cas9 technique has become the gold standard to study gene function in animal models (187).

3 Aims of the study

Over the past decade, advances in the field of genetics have led to identification of huge number of genetic variants that affect complex traits. Through GWAS, our group has been involved in the discovery of common variants that associate with changes in pubertal timing in the healthy population. However, understanding the causal biological mechanisms behind these associations represents a major challenge. The overall aim of my thesis has been to explore how we can move from the GWAS results to the next level: to move from the association signals into decoding the underlying biological mechanisms. Particularly, in this thesis project I have focused on understanding the molecular mechanism by which *LIN28B* affects the timing of sexual maturation in humans and other vertebrates.

The specific aims of this thesis were:

I) To assess whether pubertal timing associated sequence variants near *LIN28B* contribute to body size and metabolic profiles in adult humans (Study I)

II) To understand how the pubertal timing associated sequence variation affects *LIN28B* expression and whether this has consequences on the function of the HPG-axis (Studies II and III)

III) To create *lin28b* knockdown, overexpression and knockout models in zebrafish 1) to assess how these affect fish growth and 2) to explore the fundamental molecular mechanisms by which *LIN28B* might affect the timing of sexual maturation and growth in vertebrates (Studies II and III)

4. Materials and Methods

4.1 Subjects (Study I)

The study I was based on subjects from the Finnish National FINRISK study. The FINRISK studies belong to a series of Finnish population-based surveys and have been performed since 1972 at five-year intervals. The main objective of the FINRISK studies is to characterize risk factors for non-communicable disease (188). The subjects in the FINRISK cohorts have been collected from six distinct geographical regions in Finland. The cohorts have been designed to include at least 200 subjects of each sex in 10-year age groups (range 25 to 74 years) from each region. The study I included altogether data from 26,635 subjects (5,616 subjects from FINRISK 1992, 6,869 from FINRISK 1997, 8,257 from FINRISK 2002 and 5,908 subjects from FINRISK 2007). All the subjects had been genotyped. All participants had given a written informed consent of participation, and the study protocols had been approved by the institutional ethical review boards of the National Institute for Health and Welfare, and the Helsinki University Hospital, Helsinki, Finland.

4.2 Data collection and genotyping (Study I)

The participants of the FINRISK studies had been invited to a health inspection during which height, weight, waist and hip circumferences were measured by trained nurses using standardized procedures (188). For 94 and 92 subjects included in the Study I height and weight were self-reported. Before the inspection, the participants had been instructed to fast for at least four hours. The protocol included taking blood samples for 1) DNA extraction 2) determination of blood lipids and 3) assessing blood lipoprotein concentrations. For a subset of subjects from FINRISK 2007, blood glucose and insulin were measured during an oral glucose tolerance test based on ingesting 75 g of glucose. Descriptive cohort statistics are shown in **Table 3**.

The study subjects were genotyped for SNPs rs7759938 and rs314279 using a Sequenom MassARRAY genotyping platform (Sequenom Inc., San Diego, USA). Genotypes were automatically called with Typer 4.0 Analyzer software, and the clustering of the genotypes

was manually checked. For both markers, the success rate for genotyping was >98%. The minor allele frequencies (MAF) in the sample were 0.32 for rs7759938 and 0.13 for rs314279. We observed a pair-wise correlation (r2) of 0.3 between the markers. For all the participating cohorts, the genotype frequencies were in Hardy-Weinberg equilibrium at both loci.

Table 3. Descriptive statistics of the cohorts included in study I

FINRISK 1992	**Men**				**Women**			
Response Variable	n	Mean (SD)	Min	Max	n	Mean (SD)	Min	Max
Age(yrs)	2589	44.8 (11.1)	25	64	3027	44.8 (11.4)	25	64
Height(cm)	2589	176.1 (6.8)	148.0	202.0	3026	162.5 (6.1)	130.0	182.0
Weight(kg)	2589	82.5 (13.0)	46.5	154.0	3026	67.9 (12.6)	36.0	137.0
BMI(kg/m2)	2589	26.6 (3.9)	17.5	53.3	3025	25. 8 (4.9)	15.4	52.0
Waist(cm)	2587	94.1 (11.3)	64.0	144.0	3027	80.4 (11.7)	58.0	139.0
Hip(cm)	2586	101.7 (6.6)	51.0	153.0	3026	101.7 (8.7)	80.5	148.0
WHR	2586	0.92 (0.07)	0.65	1.67	3025	0.79 (0.07)	0.62	1.09
Cholesterol (mmol/l)	1692	5.8 (1.1)	2.4	10.5	2000	5.6 (1.1)	2.4	10.3
HDL (mmol/l)	1693	1.3 (0.3)	0.5	2.7	2000	1.53 (0.34)	0.5	2.7
APOA1 (g/l)	229	1.36 (0.23)	0.38	2.04	346	1.46 (0.24)	0.64	2.21
APOB (g/l)	297	1.02 (0.23)	0.44	1.90	419	0.96 (0.23)	0.49	1.80
Triglycerides (mmol/l)	1693	1.9 (1.3)	0.4	12.8	2000	1.30 (0.87)	0.2	14.9
FINRISK 1997	**Men**				**Women**			
Response Variable	n	Mean (SD)	Min	Max	n	Mean (SD)	Min	Max
Age(yrs)	3236	48.8 (13.5)	25	74	3619	47.4 (12.9	25	74
Height(cm)	3186	175.5 (7.1)	134.0	204.0	3619	162.3 (6.4)	139.0	192.0
Weight(kg)	3185	83.0 (13.2)	40.0	150.0	3609	69.3 (13.3	37.1	137.8
BMI(kg/m2)	3184	27.0 (4.0)	14.7	48.4	3608	26.4 (5.1)	14.7	49.4
Waist(cm)	3230	94.6 (11.4)	64.5	149.0	3633	81.9 (12.4)	58.0	133.5
Hip(cm)	3228	101.9 (6.8)	77.0	146.5	3619	101.8 (9.2)	69.5	159.0
WHR	3228	0.93 (0.07)	0.68	1.20	3608	0.80 (0.07)	0.61	1.38
Cholesterol (mmol/l)	2085	5.6 (1.0)	2.8	10.0	2466	5.6 (1.1)	2.6	9.7
HDL (mmol/l)	2085	1.3 (0.3)	0.4	2.8	2466	1.5 (0.4)	0.5	3.0
APOA1 (g/l)	1997	1.50 (0.26)	0.35	2.61	2349	1.72 (0.30)	0.48	2.97
APOB (g/l)	1998	1.06 (0.25)	0.21	2.09	2350	0.99 (0.25)	0.30	2.04
Triglycerides (mmol/l)	2085	1.7 (1.1)	0.3	11.4	2466	1.3 (0.8)	0.3	11.4

FINRISK 2002	**Men**				**Women**			
Response Variable	n	Mean (SD)	Min	Max	n	Mean (SD)	Min	Max
Age(yrs)	3828	48.9 (13.1)	25	74	4429	47.5 (13.2)	25	74
Height(cm)	3828	175.7 (6.8)	154.0	202.0	4426	162.4 (6.3)	110.0	196.0
Weight(kg)	3828	84.3 (14.0)	48.7	174.0	4426	69.9 (13.5)	38.3	145.4
BMI(kg/m2)	3828	27.3 (4.1)	17.0	56.0	4425	26.5 (5.1)	15.8	53.5
Waist(cm)	3824	95.5 (11.8)	59.0	158.0	4389	83.9 (12.7)	56.5	139.0
Hip(cm)	3827	98.6 (7.5)	76.5	141.0	4388	99.7 (10.1)	73.0	149.5
WHR	3824	0.97 (0.07)	0.59	1.25	4388	0.84 (0.07)	0.63	1.34
Cholesterol (mmol/l)	2510	5.8 (1.1)	2.8	10.4	3042	5.6 (1.0)	3	10.5
HDL (mmol/l)	2509	1.4 (0.4)	0.3	3.1	3038	1.7 (0.4)	0.5	3.2
APOA1 (g/l)	2507	1.45 (0.25)	0.80	2.89	3033	1.63 (0.29)	0.65	2.85
APOB (g/l)	2507	1.06 (0.25)	0.32	2.18	3033	0.96 (0.23	0.25	2.13
Triglycerides (mmol/l)	2510	1.7 (1.1)	0.3	14.2	3043	1.2 (0.7)	0.3	6.9
FINRISK 2007	**Men**				**Women**			
Response Variable	n	Mean (SD)	Min	Max	n	Mean (SD)	Min	Max
Age(yrs)	2753	51.2 (13.9)	25	74	3155	50.2 (14.0)	25	74
Height(cm)	2753	176.0 (6.9)	152.0	218.0	3155	162.8 (6.3)	137.0	185.0
Weight(kg)	2753	84.9 (14.2)	47.5	193.3	3155	71.1 (14.3)	40.1	153.0
BMI(kg/m2)	2753	27.4 (4.2)	16.0	63.3	3155	26.9 (5.4)	16.4	53.1
Waist(cm)	2748	98.1 (12.0)	66.5	173.5	3106	87.9 (13.7)	60.0	142.0
Hip(cm)	2746	100.3 (7.7)	77.0	161.0	3106	101.6 (11.0)	75.5	162.0
WHR	2746	0.98 (0.07)	0.78	1.25	3105	0.86 (0.06)	0.69	1.15
Cholesterol (mmol/l)	1751	5.4 (1.0)	2.2	9.4	2210	5.4 (1.0)	2.6	10.0
HDL (mmol/l)	1750	1.3 (0.3)	0.5	3.0	2210	1.6 (0.4)	0.7	3.1
APOA1 (g/l)	1751	1.52 (0.25)	0.68	2.66	2210	1.70 (0.29)	0.87	2.92
APOB (g/l)	1751	0.99 (0.22)	0.15	1.91	2210	0.91 (0.22)	0.31	1.93
Triglycerides (mmol/l)	1751	1.6 (1.0)	0.3	10.8	2210	1.2 (0.7)	0.3	8.4
FP Glucose (mmol/l)	1853	5.88 (0.45)	4.30	7.00	2302	5.60 (0.46)	4.24	6.98
2H Glucose (mmol/l)	1841	6.12 (1.70)	2.30	11.03	2233	6.25 (1.64)	2.07	11.09
FS Insulin (mU/l)	1880	6.58 (3.41)	2.40	25.90	2253	6.13 (3.19)	2.40	25.60

BMI = Body mass index, WHR = Waist to hip ratio, ApoA1 = Apolipoprotein A1, ApoB = Apolipoprotein B, FP = Fasting Plasma, 2H glucose = Plasma glucose concentrations 2 h after a 75g oral glucose load, FS = fasting serum

4.3 Experimental animals (Studies II and III)

Zebrafish for all experiments in studies II and III were obtained from the "Turku" strain, maintained in the Panula laboratory (Department of Anatomy and Neuroscience Center, University of Helsinki) for almost two decades (189, 190). Embryos, larvae and adult zebrafish were raised and maintained at in a 14h/10h light/dark cycle at 28 °C. Zebrafish embryos and larvae were raised in 1xE3 medium (5.00mM NaCL, 0.44mM $CaCl_2$ 0.33mM MgSO4 and 0.17mM KCl. After seven days, the fish were transferred into system water. Fish husbandry was performed in the Panula laboratory at all times. Larval fish were fed SDS dry food (SDS Diets, Essex UK) once a day. Starting from 14 dpf the feeding occurred twice a day including also artemia (Great Salt Lake Artemia Cysts, Sanders, USA). To collect samples for experiments, the larvae, juvenile and adult fish were either first killed with ice-cold water, or they were anesthetized with 0.03% tricaine and killed by dissecting the brain. The well-being of the fish was monitored regularly and neither sick, injured, nor deformed fish were used in the experiments. The experiments were conducted in the Zebrafish Unit in the University of Helsinki Neuroscience center supported by Biocenter Finland, with permission from the Animal Experiment Board of the Regional State Administrative Agency of Southern Finland (ESAVI/7434/04.10.07/2016), in agreement with the ethical guidelines of the European convention for the protection of vertebrate animals used for scientific purposes.

4.4 Microinjections to dysregulate *lin28b* (Study II)

In study II, we utilized microinjections into 1-cell stage zebrafish embryos to dysregulate *lin28b* expression transiently during the embryogenesis. The microinjections that we used contained either morpholino oligonucleotides (MO) to downregulate *lin28b* expression, or synthetic *lin28b* mRNA to increase *lin28b* expression. We designed antisense morpholino oligonucleotides (MO) to target the initiator codon and a splice donor site between exon 1 and exon 2 of *lin28b* (*lin28b* MO1, 5' GCGCGGCCCTCCTTCGGCCATGTTT'3; *lin28b* MO, 5'CCCTCCTGTGCGGTCAAGAGAGAAA'3. These and a p53-blocking MO and a standard control MO (ctrl) (p53 5'GCGCCATTGCTTTGCAAGAATTG'3, ctrl 5'CCTCTTACCTCAGTTACAATTTATA'3) were obtained from Gene Tools LLC, Philomath, OR. For the injections, we tested several

different concentrations of the morpholinos, ranging from 0.75 ng to 6 ng/embryo. The MO binding over the splicing site between *lin28b* exon1 and exon2, preventing normal mRNA processing and protein translation was used in most of the experiments. We performed extensive quality control to counteract potential off-target effects associated with MO use. First, we assessed the dose response curves of several MO doses in order to avoid using MO concentrations that would result in lethality or gross developmental defects for the majority of the embryos. Secondly, mRNA rescue experiments co-injecting MO and mRNA were utilized to assess the specificity of the observed phenotypes. The inhibition of *lin28b* mRNA in zebrafish embryos by MOs was finally confirmed by qPCR.

In order to produce the synthetic *lin28b* mRNA used in the injections, we first had to clone and characterize the full-length *lin28b* transcript from zebrafish. *lin28b* full-length open reading frame cDNA constructs were prepared by RT-PCR using Phusion High-Fidelity PCR Master Mix (Thermo Fischer Scientific, Waltham, MA). Primers F 5'GATTTCGCTGGAACTTTG'3 and R 5'CTGTTTGAGGTAGATGATTTC'3, amplifying the whole predicted open reading frame (ORF) of *lin28b* (213aa) + 126bp of 5' and 58bp of 3' UTR, were used to isolate the full-length *lin28b* coding sequence. The amplified PCR products were subsequently cloned into pGEM-T Easy vector (Promega, Madison, WI) and successful insertion of the *lin28b* fragment in to the plasmid was verified by sequencing. The insert sequence was then transferred from the pGEM-T easy into plasmid pMC (donated to Panula laboratory by Dr. Thomas Czerny). The pMC plasmid was linearized with Thermo Fisher Scientific Fast Digest NotI to allow for mRNA synthesis. mMESSAGE mMACHINE T7 kit (Ambion, Austin, TX) was subsequently used to synthesize full-length *lin28b* ORF transcripts. Finally, the resulting synthetic mRNA was injected into zebrafish embryos in concentrations ranging from 125 pg to 1 ng per embryo.

4.5 Generation of *lin28b* mutant zebrafish using CRISPR-Cas9 system (Study III)

In order to create *lin28b* knockout zebrafish, we designed two CRISPR-Cas9 models targeting exon1 and exon3 of zebrafish *lin28b*. The target site for exon1 was 5'-GGAGGGTCCGCGGGGAAACA-3' and for exon3 5'-GGCTTCAGGAGCCTGCGGGA-3'

For mutant generation, 30 zebrafish embryos (F0) were injected with 600pg Cas9 mRNA and 50pg exon1 or 25pg exon3 gRNA. The injection mixtures are described in **Table 5**. Mutations in *lin28b* were induced in approximately 80% of F0 fish, based on collecting and sequencing 5 potentially mosaic F0 embryos per sample. At 30 days of age, the remaining fish were genotyped for *lin28b* mutations from tail fin samples, and the fish showing the presence of potential mutations based on sequence analysis were grown into adulthood. Assuming that for some of these fish the mutations were present also in the germline cells, the F0 zebrafish were mated back into the Turku strain to obtain zebrafish that might be heterozygous carriers of *lin28b* mutations. These F1 zebrafish were subsequently screened for the presence of *lin28b* mutations, and once present, they were grown and genotyped from tail fin samples. In case the fish carried an interesting mutation, these fish from the F1 generation were crossed to obtain F2 control (+/+), heterozygous (+/-) and KO (-/-) (KO) zebrafish for the analyses. Both CRISPR-Cas9 models (targeting either *lin28b* exon1 or exon3) produced mutant fish that were used to generate the data for the study. The "exon1" *lin28b* KO fish were homozygous for a 4bp deletion + one bp mismatch including deletion of the ATG-initiation codon in the full length *lin28b* transcript. The "exon3" fish were compound heterozygous for mutations that either deleted 10bp from the exon3, or inserted 10bp + 3 mismatches into the canonical zebrafish *lin28b* sequence. All the mutations were predicted to lead to generation of truncated *lin28b* proteins. The presence of the mutations in the knockout fish were confirmed by cDNA sequencing (**Figure 8**) and mutated sequences are listed in **Appendix 1**.

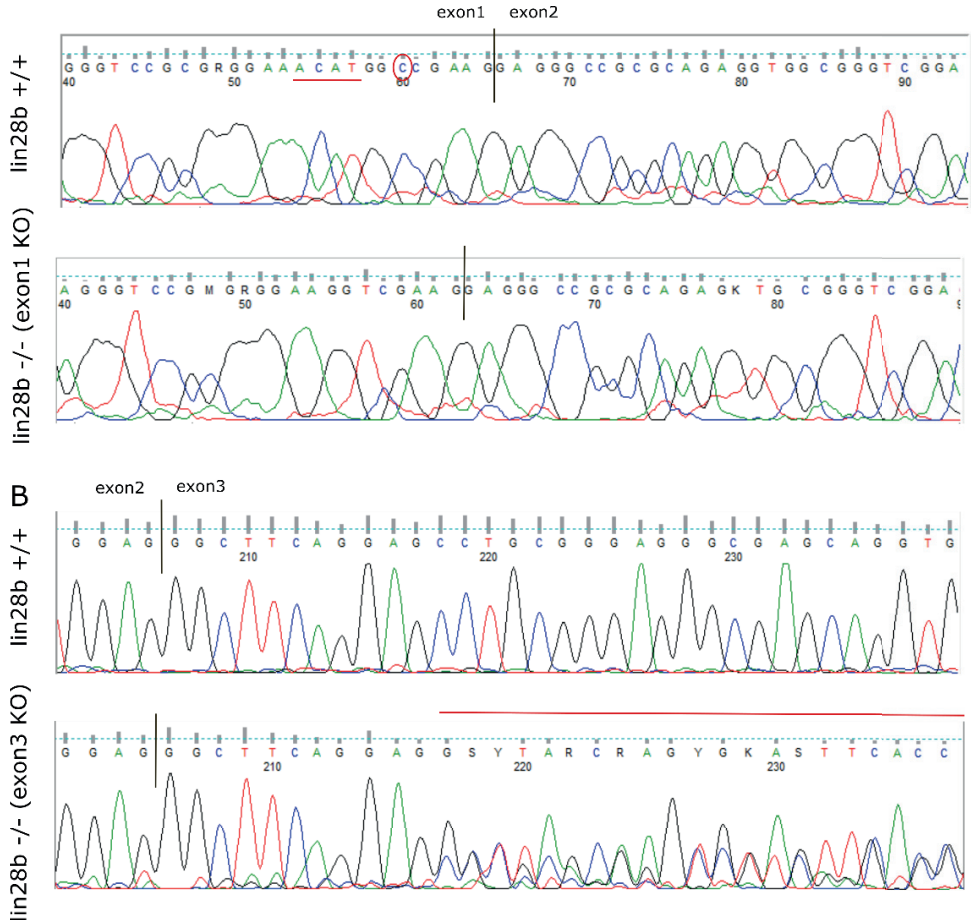


Figure 8. Sequencing traces from cDNA of *lin28b* exon1 + exon3 KO fish with control traces shown. The traces show the presence of mutations in the KO fish, leading to introduction of premature stop codons in the mRNA (data not shown). Red lines and circles indicate mutated bases, and black lines mark exon boundaries.

Appendix 1. Annotated zebrafish *lin28b* cDNA sequences showing the locations of CRISPR-Cas9 induced mutations.

A) Wild-type *lin28b* cDNA sequence. Primer locations shown in green, 5' and 3' UTR shown in lower case letters and exons in uppercase. * denotes exon boundary.

B) cDNA sequence for exon1 mutant fish, showing the deletion of four bases overlapping the translation intiation codon (ATG) and one mismatch (C>T). The only other potential intiation codon encoding for lin28b related transcript bolded and highlighted on red. The truncated protein potentially produced using this site would not contain the cold-shock and zinc-finger domains crucial for *lin28b* function.

C) cDNA sequence of exon3 mutant fish A, showing the deletion of 10 bases in *lin28b* exon 3. Premature stop codon bolded and highlighted in red.

D) cDNA sequence of exon3 mutant fish B, showing the insertion of 10 bases (red) + 3 mismatches (orange) compared to the wild-type sequence. Premature stop codon bolded and highlighted in red.

A) 5'gatttcgctggaactttggaacggagggtccgcggggaaacATGGCCGAAG*GAGGGCCGCGCAGAGGTGGCGGGTCGGACACCGCCAGGACTCCGCCG
CAGAGTCTGTCTGGCTCGGGTTACTGCAAGTGGTTCAATGTCCGCATGGGGTTTGGATTTATATCGATGACCAGCAGCGAGGGGAAGCCAGTGGACCCTCCGC
TAGACGTGTTCGTTCACCAAAGTAAGCTGGTGATGGAG*GGCTTCAGGAGCCTGCGGGAGGGCGAGCAGGTGGAGTTCACCTTTAAGAGGTCGAGTAAAGGTC
TGGAGTCGCTCCGGGTGACGGGGCCCGGGGGAGGCCCCTGCTCTGGCAGCGAGCGACGCCCCAAAGCAAAGGCCCCGCCCCTCAAACGCAAACCAAAGGGAGA
CCGGTGTATAACTGTGGAGGTCTGGACCACCACGCTAAAGAGTGTGGCCTTCCACCCCAGCCAAAGAAGTGTCACTACTGTCAGAGTGTCACGCACATGGTGG
CCCAGTGTCCCCACAAGGGGGCGCCGTCGCCCTCCGCGTCTCAGGACCCGCAACGCCCCTCCACCTCCGCTCAGTCCCCGGAAGAGGAAAGCCGCTCAGGCTC
ATCTTCATCCCCGGAGGAGGCTTCTCAAAGAGGGAGTCGCTCCCAGCGCTGGAGAAAGAGCCGGGACTGAaacaaacacacctctcatccagaccaggcccgg
acacgaaatcatctacctcaaacag3'

B) 5'gatttcgctggaactttggaacggagggtccgcggggaa----GGTCGAAG*GAGGGCCGCGCAGAGGTGGCGGGT
CGGACACCGCCAGGACTCCGCCGCAGAGTCTGTCTGGCTCGGGTTACTGCAAGTGGTTCAATGTCCGCATGGGGTTTGGATTTATATCGATGACCAGCAGCGA
GGGGAAGCCAGTGGACCCTCCGCTAGACGTGTTCGTTCACCAAAGTAAGCTGGTGATGGAG*GGCTTCAGGAGCCTGCGGGAGGGCGAGCAGGTGGAGTTCAC
CTTTAAGAGGTCGAGTAAAGGTCTGGAGTCGCTCCGGGTGACGGGGCCCGGGGGAGGCCCCTGCTCTGGCAGCGAGCGACGCCCCAAAGCAAAGGCCCCGCCC
CTCAAACGCAAACCAAAGGGAGACCGGTGTATAACTGTGGAGGTCTGGACCACCACGCTAAAGAGTGTGGCCTTCCACCCCAGCCAAAGAAGTGTCACTACTG
TCAGAGTGTCACGCACATGGTGGCCCAGTGTCCCCACAAGGGGGCGCCGTCGCCCTCCGCGTCTCAGGACCCGCAACGCCCCTCCACCTCCGCTCAGTCCCCG
GAAGAGGAAAGCCGCTCAGGCTCATCTTCATCCCCGGAGGAGGCTTCTCAAAGAGGGAGTCGCTCCCAGCGCTGGAGAAAGAGCCGGGACTGAaacaaacaca
cctctcatccagaccaggcccggacacgaaatcatctacctcaaacag3'

C) 5'gatttcgctggaactttggaacggagggtccgcggggaaacATGGCCGAAG*GAGGGCCGCGCAGAGGTGGCGGGTCGGACACCGCCAGGACTCCGCCG
CAGAGTCTGTCTGGCTCGGGTTACTGCAAGTGGTTCAATGTCCGCATGGGGTTTGGATTTATATCGATGACCAGCAGCGAGGGGAAGCCAGTGGACCCTCCGC
TAGACGTGTTCGTTCACCAAAGTAAGCTGGTGATGGAG*GGCTTCAGGAG----------
GGCGAGCAGGTGGAGTTCACCTTTAAGAGGTCGAGTAAAGGTCTGGAGTCGCTC
CGGGTGACGGGGCCCGGGGGAGGCCCCTGCTCTGGCAGCGAGCGACGCCCCAAAGCAAAGGCCCCGCCCCTCAAACGCAAACCAAAGGGAGACCGGTGTATAA
CTGTGGAGGTCTGGACCACCACGCTAAAGAGTGTGGCCTTCCACCCCAGCCAAAGAAGTGTCACTACTGTCAGAGTGTCACGCACATGGTGGCCCAGTGTCCC
CACAAGGGGGCGCCGTCGCCCTCCGCGTCTCAGGACCCGCAACGCCCCTCCACCTCCGCTCAGTCCCCGGAAGAGGAAAGCCGCTCAGGCTCATCTTCATCCC
CGGAGGAGGCTTCTCAAAGAGGGAGTCGCTCCCAGCGCTGGAGAAAGAGCCGGGACTGAaacaaacacacctctcatccagaccaggcccggacacgaaatca
tctacctcaaacag3'

D) 5'gatttcgctggaactttggaacggagggtccgcggggaaacATGGCCGAAG*GAGGGCCGCGCAGAGGTGGCGGGTCGGACACCGCCAGGACTCCGCCG
CAGAGTCTGTCTGGCTCGGGTTACTGCAAGTGGTTCAATGTCCGCATGGGGTTTGGATTTATATCGATGACCAGCAGCGAGGGGAAGCCAGTGGACCCTCCGC
TAGACGTGTTCGTTCACCAAAGTAAGCTGGTGATGGAG*GGCTTCAGGAGCCTTCAGGAGCCTTCAGGAGGCGAGCAGGTGGAGTTCACCTTTAAGAGGTCG
AGTAAAGGTCTGGAGTCGCTCCGGGTGACGGGGCCCGGGGGAGGCCCCTGCTCTGGCAGCGAGCGACGCCCCAAAGCAAAGGCCCCGCCCCTCAAACGCAAAC
CAAAGGGAGACCGGTGTATAACTGTGGAGGTCTGGACCACCACGCTAAAGAGTGTGGCCTTCCACCCCAGCCAAAGAAGTGTCACTACTGTCAGAGTGTCACG
CACATGGTGGCCCAGTGTCCCCACAAGGGGGCGCCGTCGCCCTCCGCGTCTCAGGACCCGCAACGCCCCTCCACCTCCGCTCAGTCCCCGGAAGAGGAAAGCC
GCTCAGGCTCATCTTCATCCCCGGAGGAGGCTTCTCAAAGAGGGAGTCGCTCCCAGCGCTGGAGAAAGAGCCGGGACTGAaacaaacacacctctcatccaga
ccaggcccggacacgaaatcatctacctcaaacag3'

Table 4. Primers, cycling conditions and target sequences used in the generation and analysis of the *lin28b* knockout zebrafish.

Mutant strain	PCR primers for mutation sequencing	PCR cycling conditions for sequencing
exon1	F: 5'-TACAAACAACGTAAACAAAG-3'; R:5'-AAAGAGACATGACTAAATATC-3'	95°C 3min; 35x(95°C 30s, 55°C 30s, 72°C 1min); 72°C 10min
exon3	F: 5'-GACTCCAGACCTTTACTC-3'; R: 5'-TATACAAGAACGCCTGAT-3'	95°C 3min; 35x(95°C 30s, 55°C 30s, 72°C 1min); 72°C 10min
	HRM primers for mutation detection	**HRM cycling conditions**
exon1	NA	NA
exon3	F: 5'-CCTCTTAAAGGTGAACTCCA-3'; R:5'-GGGTGTAATAAATATACAAGAACG-3'	95°C 5min; 40x(95°C 30s, 55°C 30s, 72°C 30s); 72°C 10min
	Primers for *lin28b* amplification from the cDNA of the mutant fish	**PCR cycling conditions for cDNA sequencing**
exon1 and exon3	F:5'-GATTTCGCTGGAACTTTG-3'; R:5'CTGTTTGAGGTAGATGATTTC-3'	95°C 3min; 35x(95°C 20s, 55°C 20s, 72°C 30s); 72°C 10min
	guideRNA Target Sequence	
exon1	5'-GGAGGGTCCGCGGGGAAACA-3'	
exon3	5'-GGCTTCAGGAGCCTGCGGGA-3'	

Table 5. Injection mixtures to create CRISPR-Cas9 induced *lin28b* mutations to F0 zebrafish.

Compound	CRISPR-Cas9 Injection mixture *lin28b*-exon1 (194,7ng/µl)	CRISPR-Cas9 Injection mixture *lin28b*-exon3 (98ng/µl)
sgRNA exon1	0.5µl (195ng/µl)	0
sgRNA exon3	0	0.5µl (98ng/µl)
Cas9 mRNA	2µl (592ng/µl)	2µl (592ng/µl)
phenol red	0.7µl	0.7µl
H20	0.5µl	0.5µl
total	3.7µl	3.7µl

4.6 Analysis of fish growth (Studies II and III)

In study II, we characterized the growth of the zebrafish in two separate experiments. The first experiment included analyzing fish length and head circumference at 6d, and the other

examined fish growth from juveniles to puberty (60-120d). The experiments including measuring fish size were performed with the researcher taking measures blinded for the fish group. Puberty and sex were assessed by examining the growth of gonads following the guidelines from(70), with egg production determined as visible eggs upon dissecting female fish. For the second experiment comparing fish size during juvenile growth the number of fish per tank (tank size = 3 L) was kept similar between the groups (ranging between 7 vs. 7 fish to 20 vs. 21 fish). Each fish used in study II was measured only once at a single timepoint (i.e. we did not follow the growth of individual fish).

In study III, we characterized *lin28b* KO fish growth with two experiments. We first assessed larval fish size at 5d, and secondly we followed the growth of fish between 30dpf and 240dpf. In the first experiment we measured altogether 201 5-day old fish for body length and head circumference. In second experiment, we measured fish length and weight at 30d intervals (weight was measured starting from 120dpf). The fish in the second experiment were genotyped from tail fin samples either by sequencing (exon1) or by High-Resolution-Melting curve analysis (HRM) (exon3) at each time point. HRM analysis is based on different melting temperatures of PCR products overlapping the mutation site from wild-type, heterozygous and homozygous mutant samples. Therefore, it is possible to detect sample genotype in a simple real-time PCR reaction without additional sequencing steps or special reagents. The primers used in sequencing and HRM and cycling conditions are shown in **Table 4**. In both experiments, *lin28b* knockout and control fish were grown together. In the second experiment the fish were grown in three 5 liter tanks. Two of the tanks contained exon1 KOs and control fish, (starting N=32 and 34 fish per tank), and one contained exon3 KOs and controls N=38). The researcher measuring and sequencing the fish was always blinded for their genotype information.

4.7 RNA extraction, cDNA synthesis and qPCR (Studies II and III)

In studies II and III, total RNA was extracted using miRNeasy mini Kit (Qiagen Inc., Valencia, CA) according to the instructions of the manufacturer. In study II the RNA was extracted either from 20 to 25 pooled embryos/larvae collected at 1, 2, 3 and 5-dpf, or from the individual brain of 120d old fish. In study III, RNA was extracted either from 6-11 pooled 1d

or 7d samples (RNA sequencing) or from the single brain/pituitary of a 130d old fish. In study II, 0.5–1.5 µg of RNA, and in study III 0.1-0.2µg of RNA was reverse-transcribed using the SuperScript® VILO cDNA Synthesis Kit to yield the cDNA for the qPCR experiments. In study III, the 1d and 7d samples subjected to RNA sequencing were DNAse treated, showing median RNA Integrity Number (RIN) >9.5.

In both studies II and III, qPCR was performed with the Light Cycler® 480 (F. Hoffmann-La Roche Ltd, Switzerland) instrument using iQ™ SYBR® green Supermix (Bio-Rad, Hercules, CA) (Study II) or Roche Light Cycler 480 SYBR Green I master mix (Study III) as instructed (F. Hoffmann-La Roche Ltd, Switzerland). In study II, 0.5 µl of 5 mM primers and 1 µl of 1:14 dilution of the cDNA was used per 10 µl qPCR reaction, whereas in study III we utilized 0.75 µl of 5 mM primers and 2 µl of 1:10 dilution of the cDNA per 15 µl qPCR reaction Primers for amplification were designed with Sigma OligoArchitect™ and Primer-BLAST (NCBI). Primer sequences and cycling conditions are shown in **Table 6**. Fluorescence changes were monitored after every cycle. Dissociation curve analysis was performed to ensure only a single amplicon was obtained. Results were obtained with the LightCycler® 480 software and after quality control, data were normalized and calculated based on Ct values. Reactions were run up as duplicates and relative expression levels were calculated based on *b-actin* (study II&III) and *eif-1a* (study II), using the 2-δδCT method. qPCR cycling conditions were as follows: 95C° / 10min; (45x) 95 C° / 30s; 60 C° / 20s; 72 C° / 30s.

Table 6. Primers used in the qPCR analyses in the studies II and III in 5'-3' orientation.

Primers used in the qPCR analysis		
Gene	Forward primer	Reverse primer
lin28b	CTGGAACTTTGGAACGGAGG	TGCGGACATTGAACCACTTG
gnrh3	CACCAATACACATAGTGAATGAGG	CCCGTCTGTCTGGAAATCTT
kiss2	TATGCCAGACCCCAAACC	AAGACTTGCGTTCGTCAGAA
notch1a	AGAGCCGGATTCAGCGGTC	TTACAGGGACGTGGAGAACAAG
ghrh	GTGCTATGCTGCTTGTTACTATC	ATACTTGACTGACGCTTTACATTG
ss1	CCAAACTCCGCCAACTTC	CTCCAGACGCACATCATC
b-actin	CGAGCAGGAGATGGGAACC	CAACGGAAACGCTCATTGC

eif-1a	CCAACTTCAACGCTCAGGTCA	CAAACTTGCAGGCGATGTGA
esr1	CAGGACCAGCCCGATTCC	TTAGGGTACATGGGTGAGAGTTTG
esr2a	CTCACAGCACGGACCCTAAAC	GGTTGTCCATCCTCCCGAAAC
ar	CACTACGGAGCCCTCACTTGCGGA	GCCCTGAACTGCTCCGACCTC
pomca	AGCTCAGTGTTGGGAAAACG	GGTAGACGGGGGTTTCATCT
oxt	GGTGTCAGCCTTGGTGAATAAT	GTTTGAGATGTAGCAGGCCG
lhb	GCAGAGACACTTACAACAGCC	AAAACCAAGCTCTGGAGCAGCC
fshb	GATGCGTGTGCTTGTTCTGG	ACTCGATCCATTGTCCAGCAT

4.8 RNA-sequencing and processing of RNA-seq data (Study III)

Quality and quantity of the extracted RNA samples were analyzed with 2100 Bioanalyzer RNA 6000 Nano Kit (Agilent, Santa Clara, CA, USA) and Qubit RNA BR kit (Thermo Fisher Scientific, Waltham, MA, USA). For genomic DNA contamination measurement, Qubit DNA BR kit (Thermo Fisher Scientific, Waltham, MA, USA) was used. mRNA libraries were prepared from 80 ng of extracted RNA with QuantSeq 3' mRNA-Seq Library Prep Kit (Lexogen GmbH, Vienna, Austria) according to user guide version 015UG009V0221. ERCC RNA spike-in mix (Life Technologies, Carlsbad, CA, USA) was added as control to each sample according to manufacturer's instructions. Libraries were quantified and pooled for sequencing using 2100 Bioanalyzer DNA High Sensitivity Kit (Agilent, Santa Clara, CA, USA). Sequencing was performed with Illumina HiSeq2500 system in HiSeq high output mode using v4 chemistry (Illumina, San Diego, CA, USA). Read length for the paired-end run was 2x101 bp and target coverage 5 M reads for each library.

Using Chipster analysis platform, the produced sequencing reads were preprocessed with FastX and PRINSEQ (191). The preprocessed reads were aligned to zebrafish genome build GRCz10 with TopHat2, and read count tables were produced by HTSeq. Differential expression (DE) analysis for RNAseq samples was performed using EdgeR in R 3.4.3 (192, 193). Gene ontology (GO) pathway analyses of the RNA-seq data were performed with Gene

Set Enrichment Analysis (GSEA) and Gene Ontology enRIchment anaLysis and visuaLizAtion (GORILLA) tools in combination with REViGO (194, 195).

4.9 Analysis of gene expression data from the GTEx database (Studies II & III)

The Genotype-Tissue Expression (GTEx) Project is a resource providing gene expression and genotype data from hundreds of individuals and thousands of samples from various tissues (169). The data production and analysis procedures are available through the GTEx Portal (http://gtexportal.org) and described in detail in (196).

In study II we queried the GTEx database to understand how the pubertal timing associated marker rs7759938 correlates with gene expression in the *LIN28B* locus. The GTEx eQTL results for pubertal timing and longitudinal growth associated SNPs rs7759938 and gene expression data [per sample gene Reads Per Kilobase Million (RPKM) values] for *LIN28B* and other genes in the locus used in this study were downloaded from the GTEx Portal. For study III, we queried the database to examine whether *LIN28B* expression correlates with the expression of key genes involved in the regulation of pubertal timing in the hypothalamus and the pituitary.

The study II utilized the GTEx Analysis V6 release, dbGaP Accession phs000424.v6.p1, including tissue samples from 554 postmortem donors (187 females, 357 males; age range 20–70). The GTEx data consisted of RNA sequencing data from 8555 tissue samples spanning 53 unique tissue types, with genotype data available for up to 449 donors. For study III, we used the most recent GTEx release (v7), dbGaP Accession phs000424.v6.p1, including tissue samples from 714 postmortem donors (244 females, 470 males; age range 20–70), and RNA sequencing data from 11688 tissue samples spanning 53 unique tissue types.

4.10 *In situ* hybridization (Studies II and III)

In situ hybridization was used to assess *lin28b*, *kiss2* and *gnrh3* expression in zebrafish embryos (Study II&IIII) larvae (Study II&III) and adult brain (Study III). For whole mount in situ hybridization (WISH), the embryos were dechorionated manually and treated with 0.03% phenylthiourea added in E3 medium to inhibit pigmentation. After sample collection, ISH was performed on 4% paraformaldehyde (PFA)-fixed embryos, 7-dpf dissected larvae brains and 210d old adult brains based on the protocol by Thisse & Thisse (197). Digoxigenin (DIG)-labeled antisense and sense RNA probes for *lin28b*, *kiss2* and *gnrh3* had been generated using the DIG RNA labeling kit (Roche Diagnostics) based on full-length *lin28b* ORF construct cloned into p-GEM-t easy vector. Probe prehybridization and hybridization steps were conducted at 63 °C for all of the riboprobes. In situ hybridization signals were detected with sheep antidigoxigenin-AP Fab fragments (1:10,000, Roche Diagnostics). Color was developed with chromogen substrates NBT and BCIP (nitro bluetetrazolium and 5-bromo-4-chloro-3-indolyl phosphate). The samples were transferred to glycerol before microscopy and imaging.

4.11 Microscopy and imaging (Studies II and III)

Images of embryos, and larval and adult brains showing location of *lin28b*, *kiss2* and *gnrh3* mRNAs after ISH (studies II&III) were obtained with Leica DM IRB inverted microscope and a DFC 480 change-coupled device camera using 5x, 10x and 20x magnifications (Leica, Wetzlar Germany). Pictures of live 3d- (Study II), 5d- (Study III) and 6d-old (Study II) fish were taken with Leica DMi1, at 5x magnification, after the fish had been drugged with 0.02-0.03% tricaine. Fish length at 5d (Study III) and 6d (Study II) were measured from these images using the Leica Application Suite 4.5 software. The 2dpf fish shown in the supplementary data of the study II were imaged with NIKON Coolpix E4500, integrated to Leica MZFLIII. Cell counting (study III) was performed manually under inspection of the samples under the microscope, with the person counting the cells being blinded for the genotype information.

4.12 Genetic association study on *LIN28B* and serum testosterone levels (Study III)

In study III, we utilised the UK biobank database containing over 500,000 individuals, aged between 40 and 69 at the time of recruitment (19). All participants signed a written informed consent for the UKBB study. We run an association analysis against serum testosterone levels for pubertal timing associated marker rs7759938, which has been shown to correlate with *LIN28B* expression levels at the hypothalamus and the pituitary. Analysis was performed using linear mixed model association testing implemented in BOLT-LMM v2.3 software (198). Sex-specific log-transformed testosterone levels were adjusted for age and BMI, inverse normalised, and the analyses were run using 10 PCs as covariates. For the analysis including females, testosterone levels were further adjusted for menopausal status. The analyses included individuals from white British ancestry only, with testosterone levels between +/- 5SD from the population mean for each sex.

4.13 Statistical analyses (Studies I, II and III)

For study I, we ran a linear regression analyses using PLINK to assess whether pubertal timing markers rs7759938 and rs314279 would show evidence of association to adult body size and metabolic traits. Analyses were run separately for each participating cohort, assuming an additive genetic model and adjusting for subject age and sex. The two tested markers are only partially correlated, ($r^2 = 0.3$, and $D' = 1$), and we additionally evaluated the independent effect at each locus by including both SNPs into the regression model. To evaluate the collinearity of the two SNPs in the regression analyses, the variants' variance inflation factor (VIF) was computed for each anthropometric trait with R (http://www.r-project.org/) version 2.12.2 (193). The meta-analysis was done with MetABEL (http://www.genabel.org/packages/MetABEL). Additionally, multiple linear regression analysis was performed to assess SNP by age and SNP by sex interactions for both markers. To account for multiple testing, we set a conservative significance threshold ($p<0.005$), based on Bonferroni-correction accounting for ten independent tests.

The data analysis for study II was performed with R version 3.0.2. The data from the qPCR analysis including 1- 5 d old fish (showing unequal variance between the groups) was analyzed with Welch ANOVA combined with a Games-Howell post-hoc test. The analysis of qPCR results from the 120d old fish brain was done with one-way ANOVA combined with post-hoc Tukey HSD. Fish length and head circumference at 6d was analyzed using the same methods. The juvenile growth data comparing WT and mRNA injected fish was assessed by a logistic regression model adjusting for fish age. Based on the data from Chen and Ge (70) and our own growth data, we made the assumption that the fish growth was approximately linear between two and four months (60-120d). We further performed time point and sex specific analyses using Welch two sample t-tests to assess the effect on growth in more detail. To compare the sex/egg ratios in adult fish and phenotype ratios in MO and mRNA injected embryos, we utilized Fisher's exact test and Bonferroni correction for multiple testing when necessary.

In study II, a set of 20–25 embryos collected either at 1,2,3 or 5d and subjected to qPCR denotes an experimental unit. In the qPCR experiments using 120d fish, an experimental unit is a pooled sample consisting of 7-8 fish brain. Regarding the experiments assessing growth, either a well containing a single 6d larva, or a tank containing 7 to 21 juvenile fish was considered as an experimental unit. Fish growth was not tracked i.e. all the fish used in the study were measured only once.

For study III, all statistical analyses were performed with R version 3.4.3, serving as a proxy to estimate the time that preparing the work for the thesis has taken. Fish genotype ratios from the heterozygote crosses were assessed with Chi-square test. Statistical analysis assessing fish size was made with ANOVA using post-hoc Tukey HSD. The qPCR results based on 130d fish brain and pituitaries were analysed using one-way ANOVA with post-hoc Tukey HSD.

5 Results and Discussion

This chapter summarizes the results from the three original studies aiming to elucidate how *LIN28B* affects pubertal timing, growth and adult health. The results are based on combination of data sources, ranging from Finnish population cohorts to public gene expression data and to in-house created zebrafish models, and coming back to analysis of genetic data from subjects from the UK biobank. The variation in the data sources and the analysis methods characterizing this thesis reflects the complexity of the core problem we sought to unravel: how to proceed from genetic associations into understanding the related biology.

The results have been divided into six sections, based on research aims of the original studies. The first section describes and elaborates the findings from the original study I, the second and third are based on study II, and in the 4th, 5th and 6th sections summarize the results from study III. Finally, I discuss the results collectively in the chapter **5.7**.

5.1 *LIN28B* associates with several features related to body composition in humans (Study I)

The risk for many adult-onset complex diseases is known to correlate with growth patterns during fetal development and childhood. Epidemiologically, both early and late pubertal timing have been shown to correlate with the risk for many disease. These include cardiovascular disease and type 2 diabetes (T2D), which again are closely related to obesity. At the time of the first study, although the mice models had suggested *Lin28*-genes may regulate energy homeostasis and glucose metabolism, *LIN28B* had not been associated with metabolism and obesity in humans. Therefore, in the first study belonging to this thesis, we sought to assess whether the sequence variants in the *LIN28B* locus that had been associated with pubertal timing could be linked with body size and metabolic parameters in adult humans. To this end, we performed a candidate gene association study, utilizing data from the national FINRISK surveys including 26,635 Finnish adults that were subsequently genotyped for markers rs7759938 and rs314279.

5.1.1 *LIN28B* has pleiotropic effects on adult body size

In study I, we linked pubertal timing-associate sequence variation near *LIN28B* with several dimensions of human body size. First, the study replicated the previously shown associations with height. In the combined analysis including both sexes, the pubertal timing increasing allele (C) of rs7759938 was robustly associated with increased height, (β=0.042, SE=0.009, P=2x10^{-6}), corresponding roughly to 0.3 cm increase in height per allele. Notably, the effect was stronger in females (**Table 7**). Moreover, when including the other marker (rs314279) in to the regression model, the same allele of rs7759938 was associated with increased waist-to hip (WHR) ratio in males (β=0.043, SE=0.015, P=0.004).

The effects of rs314279 seemed partially overlapping to rs7759938. Similarly to rs7759938, the pubertal timing increasing allele at rs314279 (C) was associated with height (β=0.039, SE=0.012, P=0.001). Additionally, in the analysis including both sexes the same allele was associated with increased weight and hip circumference (β=0.042, SE=0.013, P=0.001, and β=0.038, SE=0.012, P=0.003, respectively). However, when conditioning with rs7759938, the association with height disappeared whereas the associations with weight and hip size stayed unchanged (**Table 7**). Although the associations were generally stronger in females, we did not find statistical evidence for sex-specific effects for the markers. Yet, based on the observation that for many of the phenotypes the associations differed when conditioning with the other marker, our results suggest that there might exist more than one causal sequence variant in the *LIN28B* locus.

The results thus suggested that *LIN28B* might have pleiotropic effects on human body size. First, we replicated the *LIN28B* association with height. The link between *LIN28B* and adult height was relatively strong already before the study, and the effect on height has since become confirmed in several other studies as well (14, 199, 200). The association with weight is likely explained by increased height: in fact when adding height as a covariate to the model, the associations with weight seem to disappear. Nonetheless, our study was the first to link *LIN28B* with increased hip circumference, an association that has been later confirmed in other GWAS studies (144). Although the WHR association with *LIN28B* has not been replicated in larger GWAS encompassing hundreds of thousands individuals (145, 200), these have suggested that the puberty delaying markers in the *LIN28B* locus may increase also

waist circumference. Interestingly, in these studies the effect on waist circumference however seems slightly bigger than the effect on hip size, hinting that the WHR association observed in our study might not be spurious. Remarkably, the direction of the effect seems to be always the same (i.e. an allele that delays puberty, likely due to increased *LIN28B* expression, always associates with an increase in a body size parameter). The associations of *LIN28B* in humans, including our study, thus closely resemble the results from various animal models where increased *LIN28*-expression has been consistently linked with larger body size, and proportional organ overgrowth (129).

5.1.2 *LIN28B* does not affect serum lipids, but may contribute to insulin levels

After assessing *LIN28Bs* contribution to body size, we wanted to evaluate whether the sequence variants rs7759938 and rs314279 might contribute to metabolic profiles in adults. ~17,500 subjects from the FINRISK cohorts included into our study had serum lipid data available, whereas data for glucose and insulin levels were available from ~4000 subjects. We did not detect any significant associations between the sequence variants near *LIN28B* and lipid/lipoprotein levels. Captivatingly, the pubertal timing increasing allele C at rs314279, was linked with lower fasting insulin levels in males (β=-0.152, SE=0.051, *P*=0.003, **Table 8**). In theory this might reflect increased insulin sensitivity, like associated with *LIN28A* and *LIN28B* overexpression in mice (130). However, since insulin data was only available for a limited number of study subjects, and up to date the result has not been replicated, it needs to be interpreted with caution.

Table 7. Association of genetic variants rs7759938 (upper panel) and rs314279 (lower panel) with anthropometric traits.

		ALL		MALES		FEMALES	
rs7759938	**N (M,F)**	**BETA (SE)**	**P**	**BETA (SE)**	**P**	**BETA (SE)**	**P**
Height (A)	26379 (12258, 14121)	**0.042 (0.009)**	**$2x10^{-6}$**	0.032 (0.013)	0.01	**0.051 (0.012)**	**0.00002**
Height (B)	26237 (12187, 14050)	**0.038 (0.010)**	**0.0002**	0.029 (0.015)	0.06	**0.046 (0.014)**	**0.001**
Weight (A)	26377 (12257, 14120)	0.015 (0.009)	0.09	0.016 (0.014)	0.25	0.015 (0.013)	0.22
Weight (B)	26235 (12186, 14049)	-0.001 (0.011)	0.93	0.009 (0.016)	0.56	-0.010 (0.015)	0.52
BMI (A)	26375 (12256, 14119)	-0.002 (0.009)	0.78	-0.001 (0.013)	0.97	-0.004 (0.012)	0.73
BMI (B)	26233 (12185, 14048)	-0.018 (0.010)	0.10	-0.008 (0.016)	0.63	-0.026 (0.015)	0.07
Waist (A)	26314 (12291, 14023)	0.013 (0.009)	0.13	0.023 (0.013)	0.08	0.005 (0.012)	0.67
Waist (B)	26173 (12220, 13953)	0.007 (0.010)	0.52	0.026 (0.015)	0.09	-0.010 (0.014)	0.48
Hip (A)	26313 (12289, 14024)	0.009 (0.009)	0.33	0.010 (0.013)	0.45	0.007 (0.012)	0.54
Hip (B)	26172 (12218, 13954)	-0.008 (0.010)	0.46	0.001 (0.016)	0.96	-0.015 (0.014)	0.29
WHR (A)	26307 (12286, 14021)	0.013 (0.009)	0.14	0.026 (0.012)	0.04	0.000 (0.012)	0.98
WHR (B)	26166 (12215, 13951)	0.021 (0.010)	0.05	**0.043 (0.015)**	**0.004**	-0.000 (0.014)	0.98
rs314279	**N (M,F)**	**BETA (SE)**	**P**	**BETA (SE)**	**P**	**BETA (SE)**	**P**
Height (A)	26288 (12213, 14075)	**0.039 (0.012)**	**0.001**	0.030 (0.018)	0.09	**0.047 (0.017)**	**0.005**
Height (B)	26237 (12187, 14050)	0.010 (0.015)	0.51	0.008 (0.022)	0.70	0.011 (0.020)	0.60
Weight (A)	26286 (12212, 14074)	**0.042 (0.013)**	**0.001**	0.025 (0.019)	0.20	**0.057 (0.018)**	**0.001**
Weight (B)	26235 (12186, 14049)	0.043 (0.016)	0.006	0.020 (0.023)	0.39	**0.063 (0.021)**	**0.003**
BMI (A)	26284 (12211, 14073)	0.026 (0.013)	0.04	0.011 (0.019)	0.55	0.039 (0.017)	0.02
BMI (B)	26233 (12185, 14048)	0.040 (0.015)	0.008	0.017 (0.022)	0.44	**0.059 (0.020)**	**0.004**
Waist (A)	26224 (12246, 13978)	0.022 (0.012)	0.08	0.011 (0.018)	0.55	0.031 (0.017)	0.07
Waist (B)	26173 (12220, 13953)	0.017 (0.014)	0.24	-0.007 (0.021)	0.76	0.038 (0.020)	0.06
Hip (A)	26223 (12244, 13979)	**0.038 (0.012)**	**0.003**	0.026 (0.019)	0.17	0.048 (0.017)	0.006
Hip (B)	26172 (12218, 13954)	**0.045 (0.015)**	**0.003**	0.028 (0.023)	0.21	**0.060 (0.021)**	**0.004**
WHR (A)	26217 (12241, 13976)	-0.003 (0.012)	0.80	-0.009 (0.017)	0.59	0.003 (0.017)	0.86
WHR (B)	26166 (12215, 13951)	-0.019 (0.014)	0.18	-0.042 (0.021)	0.04	0.003 (0.021)	0.89

All phenotypes were standardized prior to analyses. Two regression models were used: Model A including age as a covariate, Model B including both the alternative SNP and age as covariates. The effect allele for both rs7759938 and rs314279 is C. BMI = body mass index, WHR = waist to hip ratio. Table adapted from Leinonen et al. 2012.

Table 8. Association of genetic variants rs7759938 (upper panel) and rs314279 (lower panel) with metabolites

		ALL		MALES		FEMALES	
rs7759938	**N (M,F)**	**BETA (SE)**	**P**	**BETA (SE)**	**P**	**BETA (SE)**	**P**
Cholesterol (A)	17622 (7974, 9648)	-0.020 (0.011)	0.06	-0.027 (0.017)	0.11	-0.016 (0.014)	0.26
Cholesterol (B)	17530 (7930, 9600)	-0.017 (0.013)	0.19	-0.014 (0.020)	0.47	-0.019 (0.017)	0.26
HDL (A)	17617 (7973, 9644)	-0.001 (0.011)	0.94	-0.017 (0.017)	0.31	0.012 (0.015)	0.41
HDL (B)	17525 (7929, 9596)	-0.007 (0.014)	0.61	-0.032 (0.020)	0.11	0.014 (0.018)	0.44
APOB (A)	14443 (6496, 7947)	-0.022 (0.012)	0.07	-0.032 (0.018)	0.08	-0.014 (0.016)	0.38
APOB (B)	14368 (6459, 7909)	-0.010 (0.014)	0.49	-0.013 (0.022)	0.56	-0.008 (0.019)	0.68
FP Glucose (A)	4123 (1840, 2283)	0.007 (0.023)	0.74	0.015 (0.035)	0.67	0.002 (0.030)	0.95
FP Glucose (B)	4111 (1833, 2278)	-0.007 (0.027)	0.79	0.018 (0.040)	0.65	-0.027 (0.035)	0.45
FS Insulin# (A)	4099 (1865, 2234)	-0.013 (0.024)	0.58	0.002 (0.035)	0.96	-0.025 (0.032)	0.44
FS Insulin# (B)	4088 (1859, 2229)	-0.001 (0.028)	0.98	0.076 (0.041)	0.06	-0.066 (0.038)	0.08
rs314279	**N (M,F)**	**BETA (SE)**	**P**	**BETA (SE)**	**P**	**BETA (SE)**	**P**
Cholesterol (A)	17564 (7945, 9619)	-0.022 (0.015)	0.16	-0.041 (0.023)	0.08	-0.008 (0.020)	0.71
Cholesterol (B)	17530 (7930, 9600)	-0.009 (0.018)	0.63	-0.031 (0.028)	0.28	0.007 (0.024)	0.77
HDL (A)	17559 (7944, 9615)	0.010 (0.016)	0.51	0.014 (0.024)	0.55	0.007 (0.022)	0.74
HDL (B)	17525 (7929, 9596)	0.016 (0.019)	0.42	0.040 (0.029)	0.17	-0.005 (0.026)	0.85
APOB (A)	14400 (6472, 7928)	-0.038 (0.017)	0.02	-0.060 (0.026)	0.02	-0.021 (0.023)	0.34
APOB (B)	14368 (6459, 7909)	-0.030 (0.020)	0.14	-0.052 (0.031)	0.09	-0.013 (0.027)	0.63
FP Glucose (A)	4121 (1837, 2284)	0.036 (0.032)	0.27	0.003 (0.050)	0.95	0.059 (0.042)	0.16
FP Glucose (B)	4111 (1833, 2278)	0.041 (0.029)	0.16	-0.010 (0.058)	0.87	0.087 (0.050)	0.08
FS Insulin# (A)	4098 (1863, 2235)	-0.031 (0.034)	0.36	**-0.152 (0.051)**	**0.003**	0.064 (0.045)	0.15
FS Insulin# (B)	4088 (1859, 2229)	-0.029 (0.040)	0.46	**-0.211 (0.059)**	**0.0004**	0.118 (0.053)	0.03

All phenotypes were standardized prior to analyses. Two regression models were used: Model A including age as a covariate, and Model B including both the alternative SNP and age as covariates. The effect allele for both rs7759938 and rs314279 is C. ApoB = Apolipoprotein B, FP = Fasting plasma. FS = Fasting serum, 2H glucose = Plasma glucose concentrations at 2 h after a 75g oral glucose load. Participants of the oral glucose test were instructed to fast for 10 hours. # The variable was logarithm-transformed prior to the analysis. Table adapted from Leinonen et al. 20

5.2 Transient dysregulation of *lin28b* causes permanent changes in the growth of zebrafish (Study II)

After tackling *LIN28Bs* contributions to body size in humans, we wanted to develop a vertebrate model that would allow us to study the function of *LIN28B* in more detail. In study II, we concentrated on exploring whether *lin28b* dysregulation during embryogenesis might have consequences for zebrafish growth and development. Because of multiple reasons, we speculated that *lin28b* function during this period might be particularly significant. First, the gene is a stem cell factor, expressed in undifferentiated cells. In line with this concept, previous studies have suggested that *lin28b* expression is strongest during fetal development and downregulated in adults (105, 136, 201). Secondly, the primordia for all tissues, including the hypothalamus and the pituitary form already during embryogenesis. Any genetic lesions that affect normal tissue development will likely have far-reaching consequences: for example failure in the migration of the GnRH neurons to the hypothalamus at this stage can lead to absence of puberty. Finally, supporting the hypothesis that embryogenesis partly sets the frames for adult health, epidemiological data have long implied that intrauterine insults at defined periods of development may have lasting effects on physiology and metabolism (202).

Specifically, in study II we concentrated on examining how transient *lin28b* dysregulation affects zebrafish development. The precise questions we wanted to address with study II were 1) whether *lin28b* might contribute to the formation of tissues that regulate pubertal onset, and 2) whether embryonic dysregulation of the gene is enough to include changes in zebrafish growth. Hence, we mapped *lin28b* expression during the fetal stages and developed models to transiently dysregulate the gene. In these studies, we used anti-sense MO injections to knockdown the gene, synthetic mRNA injections to overexpress the gene, and combined injections of MO and mRNA to control for off-target effects of the MOs. The next sections will describe the results from these experiments, revealing how *lin28b* might contribute to the development of the hypothalamus and how embryonic *lin28b* overexpression can affect the adult size of the zebrafish.

5.2.1 *lin28b* expression peaks during embryogenesis and locates to HP axis in zebrafish larvae

We first began by assessing the expression patterns of *lin28b* during embryogenesis. Prior to our study it was already known that the *LIN28* genes are prominently expressed during early embryonic development in vertebrates, becoming largely downregulated in adult organisms (105, 136, 201). Using *in situ* hybridization to localize *lin28b* expression in zebrafish embryos and larval brain, we observed how the expression pattern of the gene evolved rapidly, reflecting the extensive tissue remodeling taking place. Consistent with the presumed function of *lin28*-genes as factors expressed in undifferentiated cells, we saw how after ubiquitous expression during the embryogenesis, the expression of *lin28b* localized to specific organ primordia in zebrafish embryos at 2dpf (**Figure 9B**). Importantly, the gene was also expressed in the brain during later development. At the time of embryonic to larval transition (3dpf), *lin28b* expression was clearly visible in the neuronal proliferation zone, but already at 7dpf the expression had become localized into a specific neuronal network that had projections to the HP axis (**Figure 9C**). Based on the expression pattern of the gene, it thus appeared that *lin28b* is widely expressed in many tissues, including neural tissues like the hypothalamus in early zebrafish development.

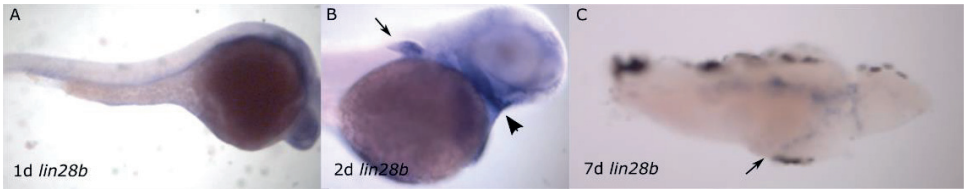


Figure 9. Visualisation of *lin28b* expression by *in situ* hybridization. Images show *lin28b* mRNA expression localisation in zebrafish embryos at 1dpf (A) and 2dpf (B) and larval brain at 7dpf (C). At first, *lin28b* is expressed throughout the embryo, but quickly becomes localized to developing organ primordia. At 7dpf, the expression of *lin28b* is already restricted to a specific neuronal network including projections into the hypothalamus. Arrow in B) = pectoral fin bud, arrowhead = pharyngeal arches. Arrow in C) = hypothalamic/pituitary region. Figure adapted from Leinonen et al. 2018 (203).

5.2.2 *lin28b* knockdown in zebrafish embryos

The analysis of *lin28b* expression in zebrafish embryos and larvae suggested that interfering with *lin28b* expression could in theory have widespread consequences for fish growth and ontogeny. Next, we sought to study how transient knockdown of the gene would affect the development of zebrafish embryos. Potentially due to the infamous off-target effects of the MOs (discussed in section 2.4.7), injecting high doses of MO (3ng and 6ng) into zebrafish embryos appeared lethal. In contrast, the smallest MO doses induced no visible changes in most fish. After determining the dose-response curves for MO injections, we chose to work with doses that caused no gross morphological defects for the majority of the fish. With these doses (0.75ng and 1.5ng), most embryos injected with *lin28b* blocking MO remained WT-like, while some still showed distinct morphological phenotypes that could not be verified to result specifically from *lin28b* knockdown (**Figure 10**). These phenotypes included small heads and defects in body axis like curved tails that have been reported as potential side effects of MOs. Although we saw some evidence for *lin28b* knockdown reducing the size of the fish, because of the uncertainty to what extent this might be caused by off target effects we refrained from detailed study of the MO injected fish growth. Instead, we concentrated on studying the development of the hypothalamus, and aimed to control the findings by including a control group that received a combined dose of MO and synthetic *lin28b* mRNA to "rescue" the embryos.

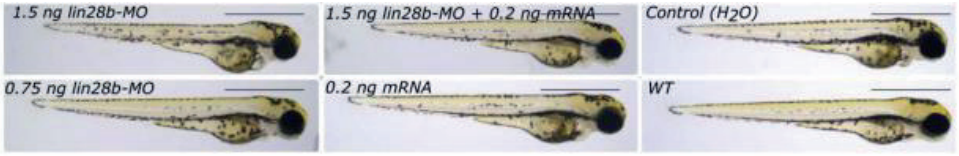


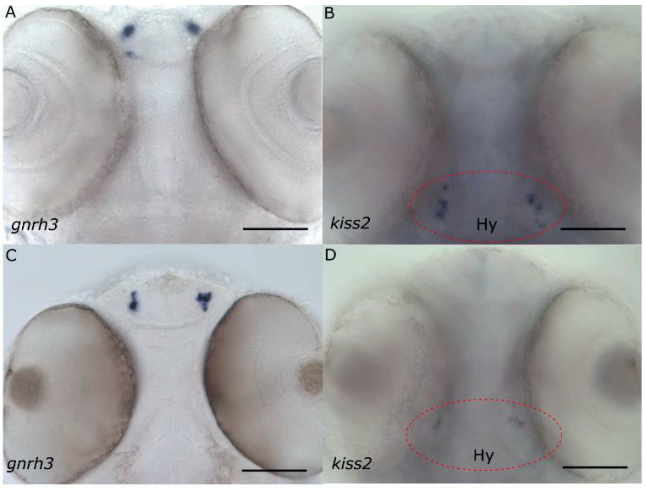
Figure 10. Representative images of 3-day old zebrafish, including fish injected with *lin28b*-blocking morpholino and synthetic *lin28b* mRNA. In all injected groups, most fish receiving injections as embryos showed no gross phenotypic abnormalities. Control injections shown here included phenol red and H20 only. Scale bars = 1 mm.

5.2.3 *lin28b* knockdown does not affect the development of Gnrh3 neurons but may disturb the emergence of Kiss2 neurons

To evaluate the potential consequences of *lin28b* knockdown on the development of the hypothalamus, we focused on studying the expression of *gnrh3* and *kiss2* in the morphants. These genes are homologous to human *GnRH* and *KISS1,* encoding for proteins which ultimately trigger pubertal onset. During the embryogenesis the expression of *gnrh3* is located at the olfactory placode, from where the *gnrh3* expressing neurons are subsequently thought to migrate to the hypothalamus (31). Conflicting one of our primary hypotheses (that *lin28b* might affect the development of the Gnrh3 neurons), knocking out *lin28b* did not seem affect the expression of *gnrh3* mRNA (**Figure 11E**). Interestingly, even in those embryos that showed mild morphological defects the *gnrh3* expression localization and intensity appeared normal (**Figure 11C**). We concluded that *lin28b* does not affect *gnrh3* expression during embryogenesis, ruling out the hypothesis that it specifically interferes with *gnrh3* neuron migration and development. Contrasting the *gnrh3* result, the expression of *kiss2* however appeared downregulated in the MO injected embryos during the first 5 days of development in qPCR analysis (P=0.033). Our result therefore suggests that *lin28b* may contribute to the development of hypothalamic Kiss2 neurons.

One critical feature distinguishes *kiss2* from *gnrh3* at this stage. Whereas the Gnrh3 neurons originate from the olfactory bulb, the Kiss2 neurons instead emerge in the hypothalamus, suggesting that *lin28b* may contribute specifically to hypothalamic development. The interpretation of the result still requires great care, especially in terms of whether these might be relevant for explaining the *LIN28B* association with pubertal timing. In fact, we do not know if *LIN28Bs* association with pubertal timing might be related to these changes at all. First, we characterized the effects of *lin28bs* during embryogenesis, and not at the time of pubertal onset or in adults. Secondly, our model does not mimic the expression changes seen in humans. Based on the GTEx data, each pubertal timing advancing allele is associated with a ~0.5 fold reduction in *LIN28B* expression, whereas in our zebrafish knockdown model we observed ~10 fold decrease in *lin28b* RNA by qPCR at 1dpf (**Figures 8 and 11**) (169). Moreover, the direction of the effect was somewhat counterintuitive: *kiss2* repression at the time of puberty would likely cause a delay in pubertal onset based on the findings indicating that mutations in *KISS1R* associate with absent puberty. For these reasons, the finding

requires replication and further study in other models, and cautiously, we interpreted the result only as a sign that *lin28b* may contribute to the formation of the hypothalamus. Interestingly, the finding that *lin28b* knockdown may interfere with *kiss2* expression nevertheless resembles the observation from the GTEx database that *LIN28B* expression correlates positively with *KISS1* expression in the hypothalamus (study III).



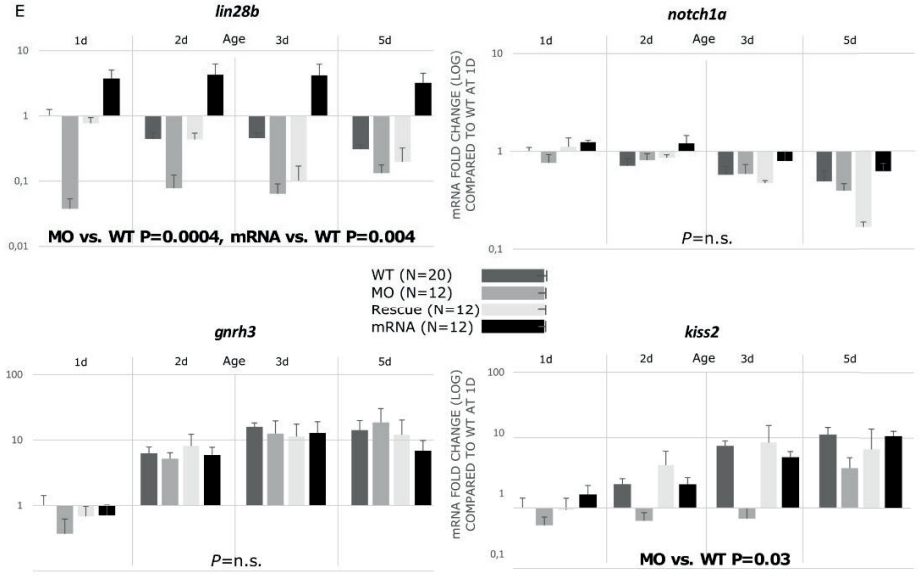


Figure 11. A-E) Visualisation of *gnrh3* and *kiss2* expression in zebrafish embryos at 3dpf and qPCR results from 1-5dpf zebrafish. A,B) = control fish, C,D) = *lin28b* MO injected fish. Note the *gnrh3* localization to olfactory bulb and *kiss2* localization at the hypothalamus. Scale bars = 100 µm. E) qPCR results from 3dpf zebrafish assessing the effect of *lin28b* up- and downregulation on selected genes during the first five days of development. Upper left panel: results show upregulation of *lin28b* expression upon synthetic *lin28b* mRNA injection (mRNA) (P = 0.00039) and downregulation of *lin28b* upon MO injection (MO) (P = 0.0042). Co-injection of *lin28b* mRNA and MO seems to result in partial rescue of *lin28b* expression (Rescue). Over the five-day period, the MO injected embryos showed reduced and mRNA injected embryos increased expression of *lin28b* compared to WT and. No significant differences were observed in the expression of *notch1a* and *gnrh3*, however *kiss2* mRNA levels were significantly reduced in the MO group (P = 0.033, Welch ANOVA with Games-Howell post-hoc test, error bars = SEM). N = 12–20. Figure adapted from Leinonen et al. 2019.

5.2.4 *lin28b* overexpression associates with increased fish growth and earlier sexual maturation

Similarly to the MO experiments, injecting large doses (1µg) of synthetic *lin28b* mRNA caused lethality and malformations for zebrafish embryos. Contrastingly, a modest mRNA dose (125ng/embryo), causing ~4-fold increase in *lin28b* expression in qPCR analysis caused no visible changes during embryogenesis. However, already by 6d post fertilization, the fish injected with synthetic *lin28b* mRNA appeared longer (P = 0.0011) than the control fish (**Figure 12A**). Similarly, their head width was also increased (P = 0.00069), indicating that receiving an extra dose of *lin28b* mRNA may enhance zebrafish growth (**Figure 12B**).

To see whether this effect on growth would persist to adulthood, the growth of the mRNA injected fish was tracked into adulthood. Remarkably, we observed that the mRNA injected fish remained on average longer than their siblings throughout juvenile development (P = 0.012, **Figure 13A**). The effect on growth was seen for fish of both sexes, although it appeared stronger in female zebrafish. Interestingly, the group receiving mRNA injections contained more sexually mature female fish (P=0.037) than the control group (**Figure 13C**). In study II we thus also reported the surprising finding that besides increasing growth, injection of synthetic *lin28b* mRNA may also advance the tempo of zebrafish growth.

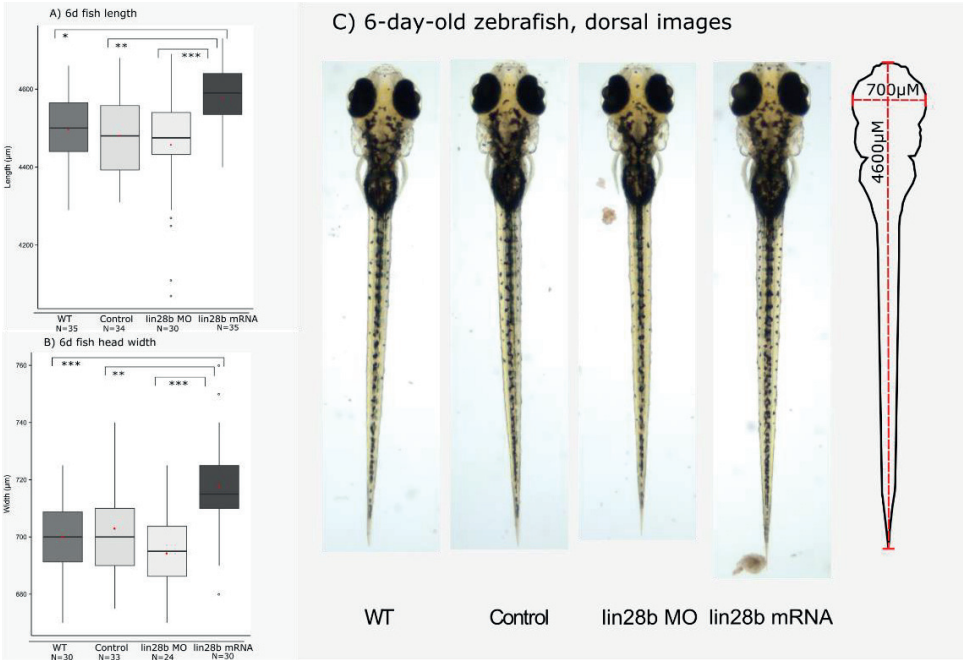


Figure 12. ***lin28b*** **overexpression increases zebrafish larval size**. Both body length and head width were increased in the group receiving mRNA injections compared to controls at 6d. Reprinted with permission from Leinonen et al. 2019.

The first result, increased size after a single injection of *lin28b* mRNA into 1-cell stage eggs, was in some ways an expected finding. The direction of the effect fits the previous data suggesting that increased *LIN28* expression is generally associated with bigger size both in humans and mice. Although we lack specific data about mRNA turnover rate, and its diffusion into different tissues, it is generally accepted that the injected mRNA stays potent during the first couple of days during development (186). More specifically, our result strongly suggests that particularly the embryonic *lin28b* expression might be crucial for determining the limits for vertebrate growth. The result is complimentary to a mice study that has suggested that fetal *Lin28B* knockout might be enough to cause reduced adult size, but *Lin28b* knockout after this stage has limited effects on mice size (131). In a wider context, these results highlight the importance of embryonic development in defining the limits for growth.

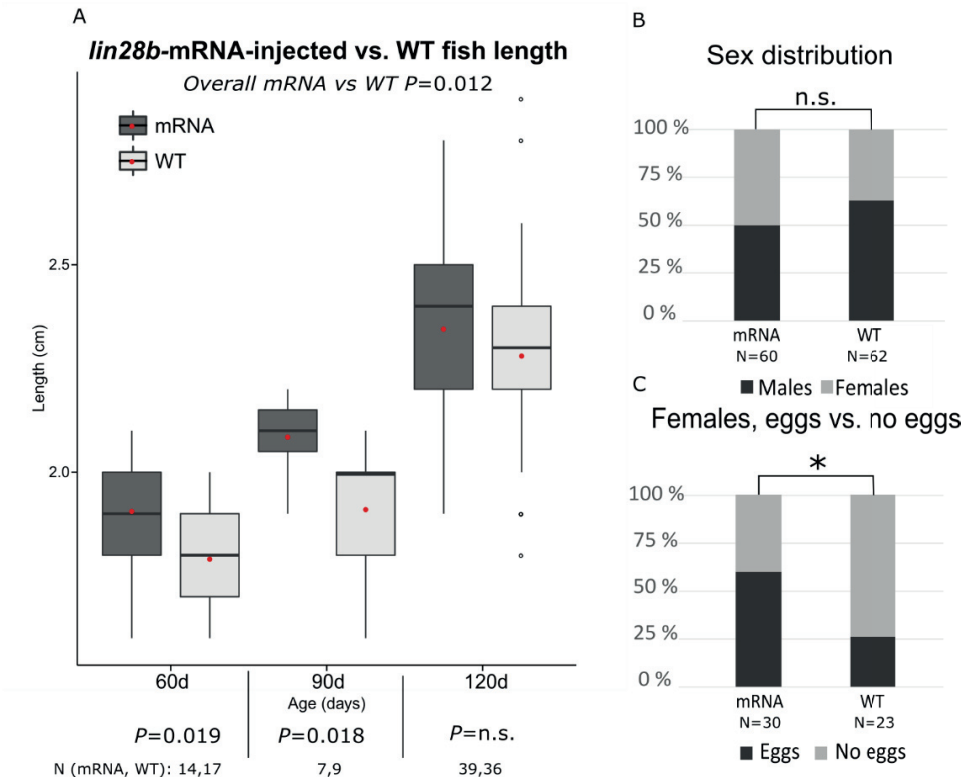


Figure 13. Transient *lin28b* overexpression causes permanent changes in zebrafish growth. A) The fish receiving an extra dose of *lin28b* mRNA as embryos remained longer than controls throughout the juvenile growth period. B Sex distribution of the fish C) The mRNA injected fish also showed tendency to mature earlier: the group receiving mRNA injections included more female fish that carried eggs at the end of the experiment. Adapted from Leinonen et al. 2019.

In contrast to the finding that *lin28b* overexpression increases fish size, the second finding that it also associated with earlier maturation of female zebrafish was rather unexpected. As opposed to our result, most other models including human data have consistently associated *LIN28B* overexpression with delayed maturation (10, 131). There is one notable exception, as one study has reported *Lin28b* knockdown to cause later puberty in male mice (133). Some plausible explanations for this rather surprising result do exist. First, in zebrafish sexual maturation is strongly dependent on body size (70), like observed in many other animals including humans (16, 50). Due to the size advantage gained by the zebrafish upon extra *lin28b*, the fish might have simply reached the critical body size required for sexual maturation earlier than the control fish. Similar mechanism might explain also the delay

puberty in *lin28b* knockout male mice that remained smaller than control mice (133). Secondly, since embryonic *lin28b* overexpression did not cause the expected pubertal delay, it may be that the strongest effects that *LIN28B* has on pubertal timing are caused by the gene's actions later in life. In the discussion of the second study, we thus speculated that the molecular mechanism by which *LIN28B* may increase organismal size might be distinct from the mechanism that affects pubertal timing.

5.3 Gene expression data indicates a role for *LIN28B* at the HP-axis in humans (Study II)

Gene expression data may provide vital clues for understanding the biology behind the GWAS associations. Particularly, it may provide answers for questions including 1) what is the causal gene in a locus, 2) what is the causal tissue and 3) what is the direction of the genetic effect i.e. does an allele linked with a trait either increase or decrease gene expression. Moreover, this type of data allows for detecting patterns and correlations between different genes, and thus may provide vital biological cues into the function of the genetic variants that have been associated with a trait. For example the GTEx database, containing genotypes and RNA expression data from postmortem human tissues provides an excellent starting point to study such questions (169).

Because of the discoveries in C. elegans and the mice model linking *LIN28A* overexpression with delayed puberty and increased size, when planning the studies for this thesis we already had a good hunch that *LIN28B* might indeed be the causal gene in the GWAS locus. Similarly, we could be rather certain that these phenotypes would be related to increased *LIN28B* expression, although this had not been officially proven. However, many other questions remained completely unanswered, including what the causal tissues are and whether the expression of other genes in the locus might also contribute to growth and puberty. Additionally, we wanted explore whether the SNPs that associated with pubertal timing truly associated with increased *LIN28B* expression. To meet these aims, we utilized the GTEx database that helped us to gain vital insight into *LIN28B* function in adult humans.

5.3.1 Pubertal timing associated SNPs affect *LIN28B* expression in the pituitary and the hypothalamus (Study II)

We first concentrated on assessing the level of *LIN28B* expression across 53 adult tissues from the GTEx database. Although the relative *LIN28B* expression was generally very low across most tissues, there were a couple of exceptions to this pattern. Especially some brain subregions and the testis showed relatively high *LIN28B* expression compared to other tissues. Fascinatingly, out of all the examined brain regions (N = 14) hypothalamus and

pituitary appeared to have the highest expression levels for *LIN28B*. Among all tissues the expression peaked in the testis (**Figure 14**).

Since the hypothalamus, the pituitary, and testes are all tissues which secrete hormones that are relevant for growth and puberty, we thought that *LIN28B* expression in these tissues might be relevant for explaining the GWAS associations. Therefore, we subsequently examined whether GWAS associated genetic variants in the *LIN28B* locus, including rs7759938 might correlate with *LIN28B* expression in these tissues. We found out that both in the hypothalamus and the pituitary the genotype at locus rs7759938 correlated significantly with *LIN28B* RNA expression (P = 0.0057 for the hypothalamus and 4.16e-05 for the pituitary). *LIN28B* was also the only protein-coding gene whose expression was affected by these sequence variants, supporting it is likely the causal gene behind the GWAS signals. As expected, the same variant in rs7759938 (C) that had been linked with a delay in pubertal timing associated with increased *LIN28B* expression. Fascinatingly, neither rs7759938 nor other sequence variants were significantly related to *LIN28B* expression levels in the testis (P = 0.21). Collectively these results suggested that *LIN28B* is indeed the causal gene in the locus, and the hypothalamus and the pituitary might be the crucial tissues to study in order to understand the molecular biology behind the GWAS associations.

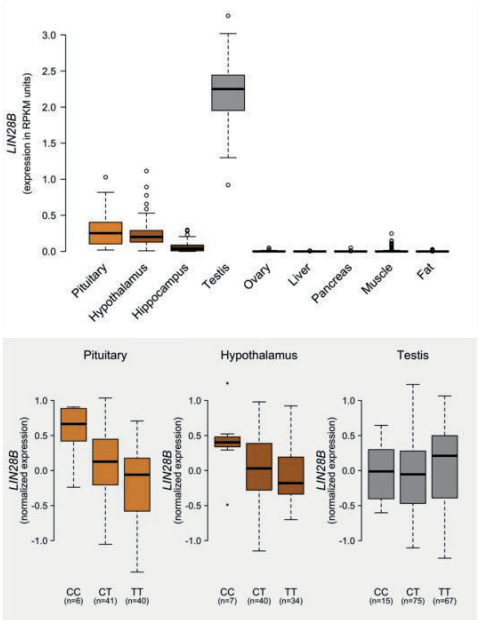


Figure 14. *LIN28B* profiling in the GTEx database. In adult humans, *LIN28B* is expressed at low levels across all tissues with the peak expression in the testis, pituitary and hypothalamus (upper panel). The pubertal timing associated SNP rs7759938 as an eQTL in the pituitary and hypothalamus, (P = 0.0057 for hypothalamus and 4.16e-05 for pituitary), but not in the testis (P = 0.21) (lower panel). The pubertal timing delaying allele (C) associates with increased *LIN28B* expression. The data shown as standardized mean expression, adjusted for principal components including sex and age. Figure reprinted with permission from Leinonen et al. 2019(203)

5.4 Insights into *lin28b* function from CRISPR-Cas9 KO models (Study III)

In contrast to the transient *lin28b* dysregulation models used in study II, we next aimed to develop a model which would allow us to study the consequences of permanent *lin28b* KO on zebrafish development. Over the past couple of years prior to our study, CRISPR-Cas9 had emerged as a powerful new technology to knockout genes (183). We chose to apply this technology in study III, preparing CRISPR-Cas9 induced *lin28b* knockout (KO) models. Because of the relative simplicity and low cost of the technology, combined with the potential to induce sequence-specific genetic lesions, several different fields of molecular biology have rapidly adapted CRISPR-Cas9 as the gold standard method to study gene function. In our case, we were able to create two different *lin28b* KO zebrafish lines that made it possible for us to study the effects permanent *lin28b* KO then has on zebrafish growth and development (**Figure 8** and section **4.6** in Methods). With our model, we specifically wanted to address two basic questions: 1) how does *lin28b* KO affect zebrafish growth and 2) whether *lin28b* KO fish would show changes in the development and the function of the HP axis, like suggested by our previous zebrafish model and the GTEx data (203).

5.4.1 *lin28b* KO affects the timing and tempo of zebrafish growth

Already before our third study, a consensus had emerged that *LIN28* genes may regulate growth in various animal species ranging from humans to nematodes: e. g. our previous study had suggested that transient, non-physiological *lin28b* overexpression can increase zebrafish body length. In the third study, we continued to assess *lin28b* effects on zebrafish size and growth by using the *lin28b* CRISPR-Cas9 KO fish we generated in-house. Upon superficial visual inspection, the *lin28b* KO fish appeared normal, and were born in the expected 1:2:1 genotypic ratio. The sex ratio of the fish was however notably biased: despite the genotype of the F2 fish, the great majority of the fish (~85-90%) were male. Hence, our results are largely based on data from male fish. After detecting no gross phenotypes in the *lin28b* KO fish upon visual inspection as embryos and larvae (**Figure 15A,** as opposed to the MO injections in study II), assessing the growth patterns of these fish became the logical

starting point for our study. Based on earlier data from humans, mice and zebrafish, our hypothesis was that *lin28b* KO zebrafish would remain smaller than control fish. We began by assessing *lin28b* KO zebrafish body size during the larval development. The development of the fish appeared generally normal, and neither head-to-tail length nor head circumference was significantly different in the KO zebrafish compared to heterozygous and wild-type siblings at 5dpf (N=201) (**Figure 15C**). On one hand the result was not surprising, since *lin28b* knockdown by MOs in study II also had non-significant effects on larval size when using non-toxic MO concentrations. Although the result did not complement the result from the mRNA injected fish showing increased growth already after six days of development, it therefore appears that downregulating *lin28b* during early development in zebrafish does not cause dramatic effects on larval size. Yet, it is worth noting that the group in which *lin28b* expression had been downregulated (MO group in study II and KO group in study III) was in both studies on average the smallest, although the result did not reach statistical significance. Hence, it remains possible that *lin28b* has subtle effects on larval size, but these may be too small to have been detected under the current models and methods.

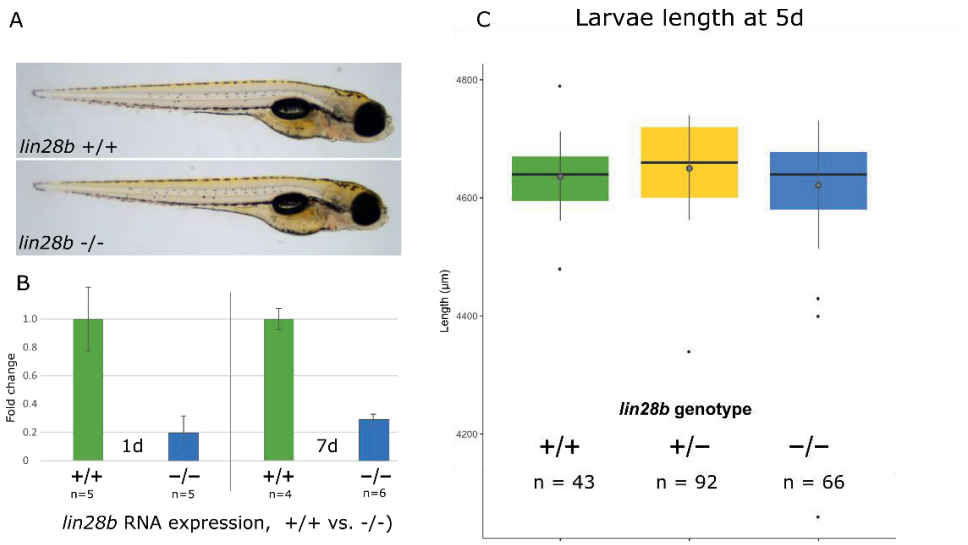


Figure 15. Characteristics of *lin28b* KO fish during larval stages. A) Images of 5d old *lin28b* control (+/+) and KO (-/-) larvae B) Full length *lin28b* mRNA expression in control vs *lin28b* KO RNA-Seq samples at 1d and 7d. C) Comparison of *lin28b* KO and control zebrafish size at 5d. No significant size differences were observed at this stage (One-Way-ANOVA with post-hoc Tukey HSD, Error bars = SEM).

We next wanted to assess whether the knockouts exhibited similar alterations in the tempo and amplitude of longitudinal growth as previously reported for humans and mice, where lower *LIN28B* expression has been linked with smaller adult size. We designed an experiment where we tracked the growth of *lin28b* KO and control fish between 30d and 240d by 30d intervals (**Figure 15A**). In this experiment, the *lin28b* KO fish appeared on average smaller than other fish already by 30dpf. The KO fish remained smaller also at the end of the experiment, matching the results from humans and mice linking lower *LIN28B* expression with smaller adult size (131, 199).

Besides the results showing that *lin28b* KO affected zebrafish size (growth amplitude), our model allowed us to observe also whether the growth proceeded in a similar fashion between the KO fish and the controls (growth tempo). Overall, zebrafish growth tempo shows extensive variation between individual fish (204). However, although the pubertal growth spurt has been thought to be somewhat specific to humans, just before sexual maturation, zebrafish show transient slowing of growth followed by a period of more rapid growth (9, 204, 205). We noticed that this transient peak in growth occurred earlier in *lin28b* KO zebrafish (**Figure 16B**). This closely resembles the finding from human data that individuals carrying genetic variants associated with lower *LIN28B* expression experience their pubertal growth spurt slightly earlier (9). Our data therefore suggests that in addition to affecting zebrafish size, *lin28b* may regulate also the tempo of fish growth in a similar manner as shown for humans. Remarkably, it appears that *LIN28B*-related genetic pathways represent a fundamental, evolutionarily conserved, mechanism to regulate vertebrate growth, despite the huge phenotypic diversity that exists for example between fish and mammals.

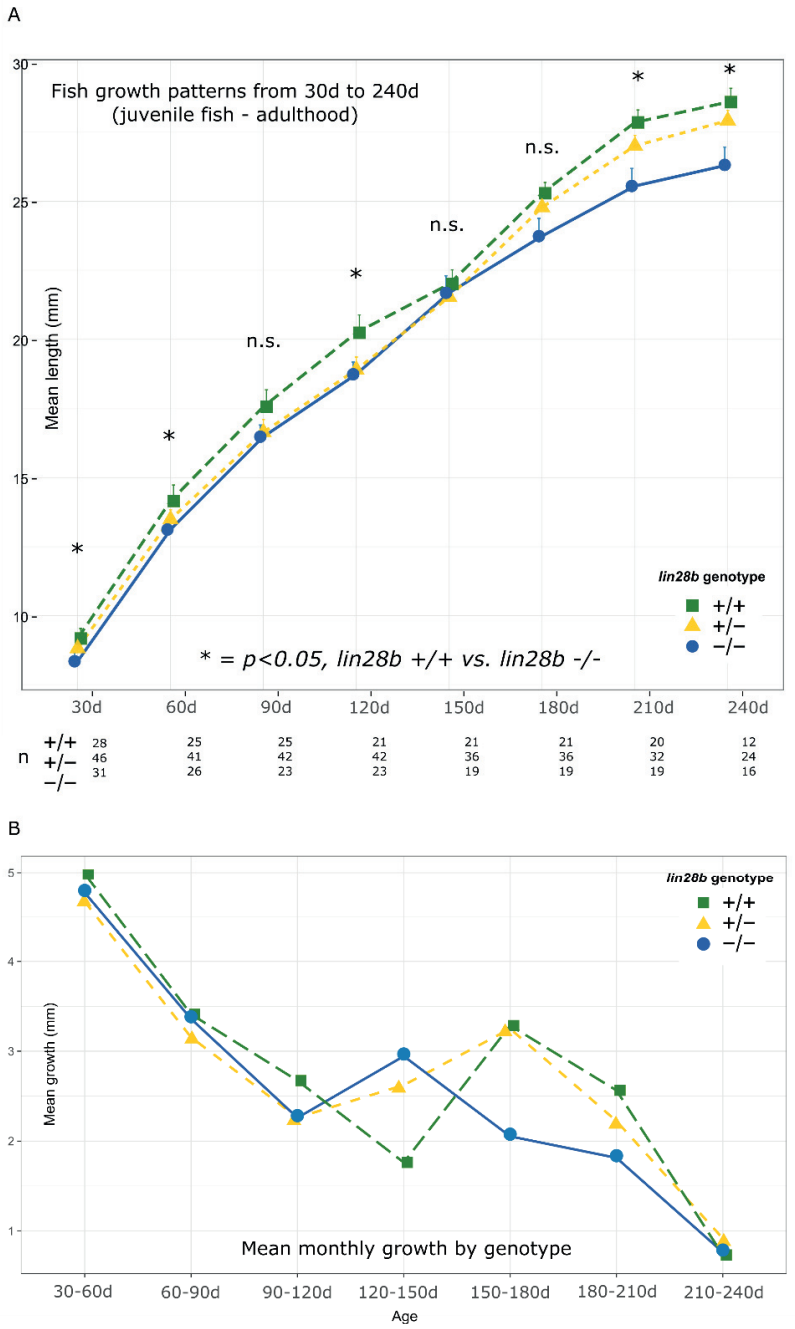


Figure 16. Characteristics and growth patterns of *lin28b* KO fish. A) Growth curves for *lin28b* KO and control fish. The KO fish remain smaller than control siblings, but the size difference temporarily vanishes around the time of sexual maturation (One-Way-ANOVA with post-hoc Tukey HSD, Error bars = SEM, *=*P*<0.05). B) Mean monthly growth by genotype. Around the period when the fish undergo sexual maturation the growth rate peaks transiently. The *lin28b* KO fish seem to undergo this transient peak in growth earlier than controls.

5.4.2 *lin28b* KO does not prevent, but may accelerate hypothalamic development in zebrafish larvae

Many of the phenotypes associated with *LIN28B* such as growth, puberty and depression are related to HP axis function, and our earlier studies had revealed that the function of the gene may be especially relevant for the function and the development of the hypothalamus both in humans and zebrafish. Since the growth patterns of the *lin28b* KO fish supported the notion that the fish may be a reliable model to study the actions by which *lin28b* affects phenotypes detected in human GWAS, we turned our focus into *lin28b* function at the hypothalamic-pituitary (HP) axis.

We first concentrated on evaluating whether *lin28b* KO fish might show visible changes in the structure and function of the HP axis. Since our previous model had suggested that *lin28b* knockdown may interfere with the development of the hypothalamus and specifically with emergence of the Kiss2 neurons, we begun by evaluating whether the knockout fish would show a reduced number, or differences in the organisation of *kiss2* expressing cells at 3dpf. However, the location of *kiss2* expression appeared similar in the KO and control zebrafish, and we observed no statistically significant differences in the number of *kiss2* expressing cells between these fish (data not shown). Although the general trend was that there were less *kiss2* cells in the KOs in the controls, we thus could not show that *lin28b* KO would block the development of the Kiss2 neurons. Potentially explaining the discrepancy between the current and our previous study (showing that MO induced *lin28b* downregulation may affect the development of the Kiss2 neurons), it is thought that gene loss by knockout may be quite well compensated by other genetic mechanisms in zebrafish. For example knocking out even key genes like *gnrh3* and *kiss2* does not lead to defects in hypothalamic development and function in zebrafish (178, 179, 206).

Despite the result that *lin28b* may not be essential for hypothalamic development in zebrafish, we continued to assess links between the gene and the HP axis. As by 7dpf the expression of lin28b in the zebrafish brain is largely restricted to specif neuronal networks including the hypothalamus and the pituitary, we next analysed RNA-seq data from 7dpf zebrafish concentrating on selected hormonal genes expressed from the HP axis. Overall, the expression patterns for most genes between the *lin28b* KO and control larvae did not differ

dramatically, but when analysing the expression of all the studied HP axis genes together, we noticed their expression was generally upregulated in the KO zebrafish at 7dpf (*p<0.05*). Out of individual genes which were analysed in the study (encoding for established hypothalamic and pituitary hormones that were expressed in high enough quantities to be observed in the RNA-seq analysis), we observed nominally significant upregulation of *chrb* and *oxt* (*p<0.005*). Moreover, in the pathway analysis of the RNA-seq data, we noted that the KO fish showed upregulation of genes that respond to estrogen signalling. During normal zebrafish development, the relative expression of hormonal genes steadily increases during the larval stages (203, 207). Altogether, the results suggest that hormonal development may be slightly advanced in the KO zebrafish at this stage, resonating with an idea that *lin28b* KO may speed up tissue maturation.

5.4.3 *lin28b* KO may affect *ESR1* expression after sexual maturation

In study III, the larval RNA-seq data encouraged checking whether we could detect changes in the expression of HP axis genes in the *lin28b* KO zebrafish. We first concentrated on assessing the expression levels of selected HP-axis genes by qPCR in 130dpf KO zebrafish, including zebrafish homologues for known puberty genes such as *gnrh3*, *kiss2*, *lhb* and *fshb*. Mimicking the human results from the GTEx database, *esr1* expression appeared downregulated in the KO fish brain compared to control fish (p<0.05) at this stage. However, no other genes were shown to be either up- or downregulated in the KO fish, including *kiss2* and *gnrh3* (**Figure 17B**). Similarly to the results from the larval fish, we observed no consistent differences in the location or staining intensity of *kiss2* and *gnrh3* expressing cells between the KOs and controls, whereas the staining pattern between individual samples in each group varied showed some variation (**Figure 17C**). Despite the hints that *LIN28B* might affect the development and function of the HP axis, excluding the potential *esr1* downregulation, we thus failed to see any dramatic changes in the *kiss2-gnrh3*-gonadotrophe-pathway in our *lin28b* KO fish after pubertal maturation. However, the *esr1* finding provided further suggestion that estrogen-related signaling might differ between the knockouts and the control fish.

5.4.4 RNA-Seq analysis links *lin28b* KO with upregulation of mitochondrial pathways in zebrafish embryos and larvae

As *LIN28B* has been associated with a wide variety of traits, besides tackling the question whether *lin28b* may play a role in the regulation of the HP axis we were also interested in screening for global changes *lin28b* KO might induce to zebrafish RNA transcriptome. Based on our earlier work we knew that the relative *lin28b* expression peaks during embryogenesis, and therefore we decided to concentrate on analyzing gene expression during embryonic and larval development. To this end, we collected samples from altogether 163 *lin28b* -/- and +/+ fish at 1dpf and 7dpf. The samples were pooled and subjected to 3' end RNA sequencing which then allowed for differential expression (DE) analyses.

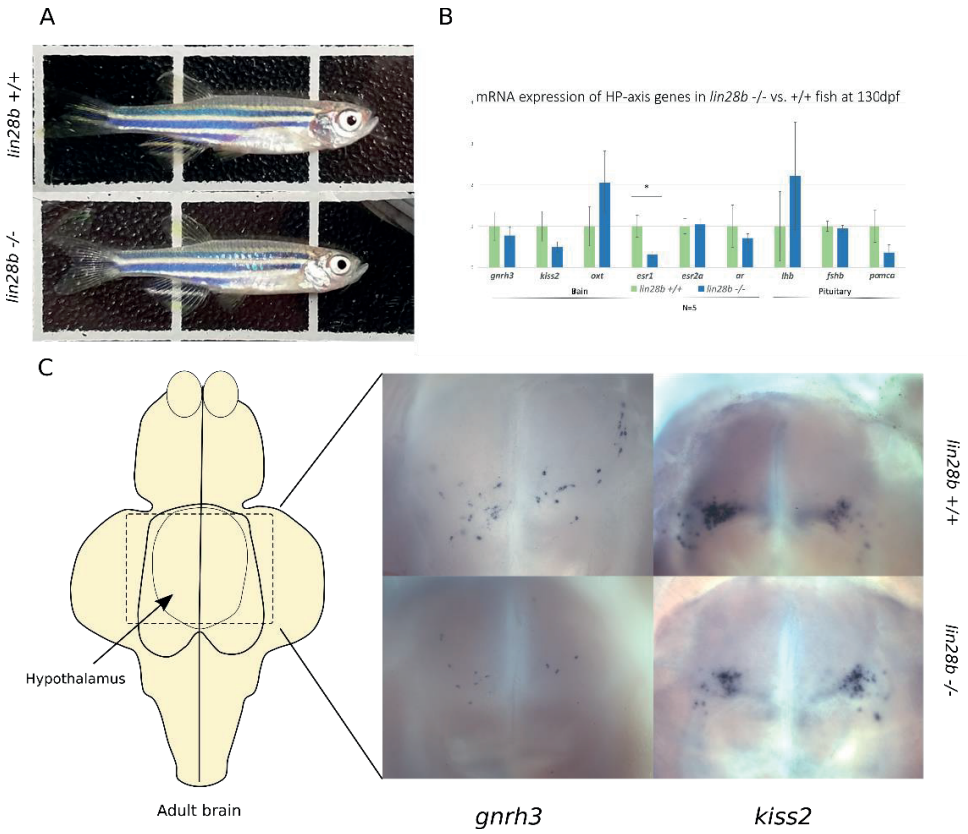


Figure 17. Characterization of HP axis development in 130dpf *lin28b* KO fish. A) images of 130dpf *lin28b* KO and control fish, also at this stage, the KO fish appeared perfectly normal and healthy, showing no distinct phenotypes when compared to the control fish. B) qPCR results from 130dpf fish showing lower *esr1* expression in the KO fish,

though in general HP axis genes show within group variation that exceeds the variation between groups. C) Visualisation of *kiss2* and *gnrh3* expression localization in the 210dpf zebrafish brain. Localization between the KO and control fish appears similar, although the number of stained cells varies between individual samples.

Owing to relatively small sample size (N=4-6 biological replicates per group), and large number of genes studied (9304-11,292) only a couple of individual transcripts passed the false discovery rate (FDR) threshold (<0.1) in the analysis (**Table 9**). We thereafter focused on identifying transcriptomic pathways that were altered in *lin28b* KO fish compared to controls. Besides the estrogen signaling finding that was discussed in section **5.4.2**, another major theme emerged from the pathway analysis of 1dpf and 7dpf fish: *lin28b* KO zebrafish showed consistently increased expression of mitochondrial genes during the first days of development (**Figure 18**). Remarkably, the result supports the earlier findings suggesting that lower *LIN28* activity promotes aerobic (mitochondrial) metabolism at the expense of glycolytic pathways: for example in murine pluripotent stem cell models loss of *LIN28* function has been associated with increase mitochondrial activity and prolonged mitochondrial mRNA half-lives (116).

We speculate that since mitochondrial metabolism is the preferred means of energy production in differentiated cells, theoretically, the result might be a reflection of *lin28b* KO zebrafish having less dividing, and more differentiated cells than controls (208, 209). This hypothesis fits the concept of *lin28b* acting as a gene that keeps cells in proliferating state. As suggested in the case of *LIN28A* and mice, eventually such tendency to prefer mitochondrial metabolism might lead to a reduction of the total number of cells in the organism, causing smaller adult size (129). Overall, the GO analyses of the global RNA-seq data thus hint that *lin28b* may contribute to the selection of cellular metabolic programs at the whole organism level in zebrafish larvae.

Table 9. Results from the differential expression (DE) analysis of RNA-seq data. Upper panel=top 15 DE genes between *lin28b* Ko and controls at 1dpf. Lower panel = top 15 DE genes at 7dpf. Positive logFC refers to genes overexpressed in *lin28b* KO zebrafish. FC=Fold change, CPM=Counts Per Million, FDR=False Discovery Rate. Gene reaching FDR<0.1 were considered to be differentially expressed between the samples.

Gene name	Gene description	logFC	logCPM	F	PValue	FDR
rrm2b	ribonucleotide reductase M2 b [Source:ZFIN;Acc:ZDB-GENE-030616-614]	1.372	3.803	38.305	0.00001	0.057
zgc:112234	zgc:112234 [Source:NCBI gene;Acc:554097]	-1.234	4.376	35.59	0.00001	0.057
apoa4b.1	apolipoprotein A-IV b, tandem duplicate 1 [Source:ZFIN;Acc:ZDB-GENE-030131-1263]	1.234	7.973	29.631	0.00016	0.378
sepp1b	selenoprotein P, plasma, 1b [Source:ZFIN;Acc:ZDB-GENE-030311-2]	0.99	5.209	21.885	0.00021	0.378
fabp2	fatty acid binding protein 2, intestinal [Source:ZFIN;Acc:ZDB-GENE-991019-5]	0.706	8.35	20.264	0.00028	0.378
si:dkey-23a13.22	si:dkey-23a13.22 [Source:ZFIN;Acc:ZDB-GENE-160113-78]	-1.223	4.396	20.973	0.0003	0.378
ces2	carboxylesterase 2 (intestine, liver) [Source:ZFIN;Acc:ZDB-GENE-061013-99]	0.776	5.7	19.734	0.00032	0.378
rnf38	ring finger protein 38 [Source:ZFIN;Acc:ZDB-GENE-030131-8693]	0.751	6.079	19.652	0.00032	0.378
bhmt	betaine-homocysteine methyltransferase [Source:ZFIN;Acc:ZDB-GENE-030131-947]	0.927	7.124	20.811	0.00038	0.385
si:ch211-150o23.2	si:ch211-150o23.2 [Source:ZFIN;Acc:ZDB-GENE-060503-386]	0.868	5.031	18.694	0.00041	0.385
CABZ01071909.3		1.216	2.725	17.813	0.00052	0.406
fga	fibrinogen alpha chain [Source:ZFIN;Acc:ZDB-GENE-031010-21]	0.857	6.216	18.31	0.00056	0.406
apom	apolipoprotein M [Source:ZFIN;Acc:ZDB-GENE-010605-5]	0.75	6.634	16.958	0.00069	0.406
lzts1	leucine zipper, putative tumor suppressor 1 [Source:ZFIN;Acc:ZDB-GENE-081104-431]	-1.252	3.018	16.513	0.00074	0.406
U6	U6 spliceosomal RNA [Source:RFAM;Acc:RF00026]	-1.127	3.393	16.153	0.00081	0.406

Gene name	Gene description	logFC	logCPM	F	PValue	FDR
zgc:153846	zgc:153846 [Source:ZFIN;Acc:ZDB-GENE-060929-898]	1.61	6.19	44.36	0.00002	0.0891
tecpr2	tectonin beta-propeller repeat containing 2 [Source:ZFIN;Acc:ZDB-GENE-060503-358]	-0.97	4.79	33.04	0.00002	0.0891
CABZ01058333.1		-1.06	7.91	34.54	0.00002	0.0891
hbae1.3	hemoglobin, alpha embryonic 1.3 [Source:ZFIN;Acc:ZDB-GENE-061207-39]	0.87	6.21	28.62	0.00005	0.1498
si:ch73-288o11.5	si:ch73-288o11.5 [Source:ZFIN;Acc:ZDB-GENE-131121-416]	0.88	4.57	25.16	0.00011	0.1995
si:dkey-247k7.2	si:dkey-247k7.2 [Source:ZFIN;Acc:ZDB-GENE-031118-45]	0.77	5.85	25.16	0.00011	0.1995
zgc:123103	zgc:123103 [Source:ZFIN;Acc:ZDB-GENE-051030-105]	1.08	4.21	23.38	0.00016	0.256
he1a	hatching enzyme 1a [Source:ZFIN;Acc:ZDB-GENE-021211-3]	1.86	4.31	29.31	0.00019	0.2617
c9	complement component 9 [Source:ZFIN;Acc:ZDB-GENE-050522-442]	0.86	6.42	21.3	0.00027	0.3406
crygm1	crystallin, gamma M1 [Source:ZFIN;Acc:ZDB-GENE-050516-5]	1.48	2.48	18.94	0.00043	0.4625
si:ch211-202a12.4	si:ch211-202a12.4 [Source:ZFIN;Acc:ZDB-GENE-030131-8486]	0.72	5.55	18.79	0.00045	0.4625
mt-nd4	NADH dehydrogenase 4, mitochondrial [Source:ZFIN;Acc:ZDB-GENE-011205-10]	0.57	9.08	17.81	0.00058	0.517
ndufa4	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 4 [Source:ZFIN;Acc:ZDB-GENE-040426-1962]	0.77	3.91	17.48	0.00063	0.517
hbae3	hemoglobin alpha embryonic-3 [Source:ZFIN;Acc:ZDB-GENE-990706-3]	0.71	5.88	17.39	0.00064	0.517
mt-nd2	NADH dehydrogenase 2, mitochondrial [Source:ZFIN;Acc:ZDB-GENE-011205-8]	0.58	9.56	17	0.00071	0.5349

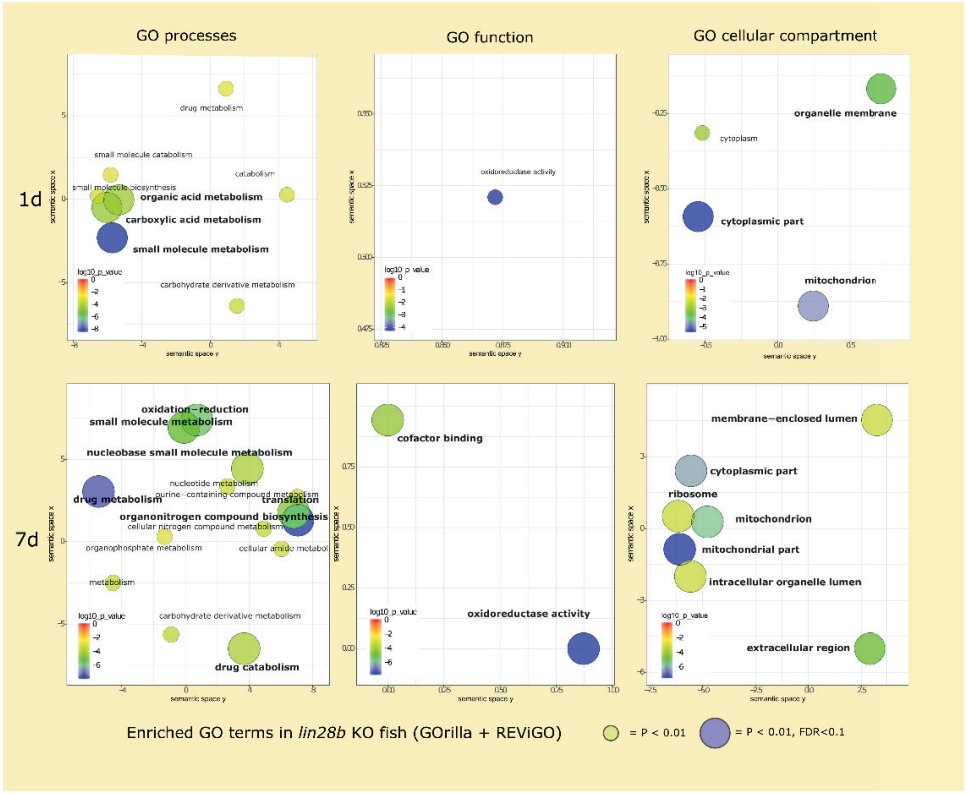


Figure 18. Summary of the GORILLA results visualized by REVIGO, indicating how top 1000 DE genes in the *lin28b* KO fish have been enriched for genes related to the function of the mitochondria.

5.5 *LIN28B* expression correlates with the expression of several HP axis genes (Study III)

Besides relying on *lin28b* KO zebrafish model, we continued to utilize the publicly available databases in study III. After showing how the pubertal timing associated sequence variation affects *LIN28B* expression at the HP axis in study II, we carried on mining the GTEx database, aiming to characterize whether *LIN28B* expression would correlate with the expression of selected HP axis genes (**Figure 19**). These genes included *ESR1, ESR2, AR, KISS1, KISS1R* and *GNRH1,* which are part of the hormonal loop that regulates sexual maturation. Unexpectedly, we observed that in adult humans, lower *LIN28B* expression seemingly correlated with lower expression of the aforementioned genes (*P*<8e-10 to <3e-19) in the hypothalamus. Surprisingly, lower *LIN28B* expression (associating with earlier pubertal onset) thus in fact seemed to correlate with lower expression of the *GNRH1* and *KISS1*, which are known to promote sexual maturation. As *KISS1* and *GNRH1* are thought to stimulate pituitary gonadotropin secretion, we expected that lower *LIN28B* levels might correlate with lower lutenizing hormone (*LH*) and follicle-stimulating hormone (*FSH*) mRNA levels in the pituitary. Contrasting this hypothesis, *LH* and *FSH* levels appeared uncorrelated with *LIN28B* expression in the GTEx dataset (**Figure 19**).

Although *LIN28B* levels did not correlate with gonadotropin mRNAs, the expression of proopiomelanocortin (*POMC*) mRNA showed a strong positive correlation with *LIN28B* expression in the pituitary (*P*<3e-17, **Figure 19**). Remarkably, *POMC* mRNA, which is a precursor for several peptides was both among the top 3 most expressed genes in the pituitary as well as among top 20 genes whose expression correlates with *LIN28B* (169). The peptides that are made based on *POMC* include b-endorphin, melanin, and adrenocorticotrophic hormone (ACTH) that regulate various physiological processes in the human body including appetite and glucocorticoid secretion. The strong association with *POMC* suggests that *LIN28B* may indeed contribute to the foundation of hormonal dynamics at the HP axis.

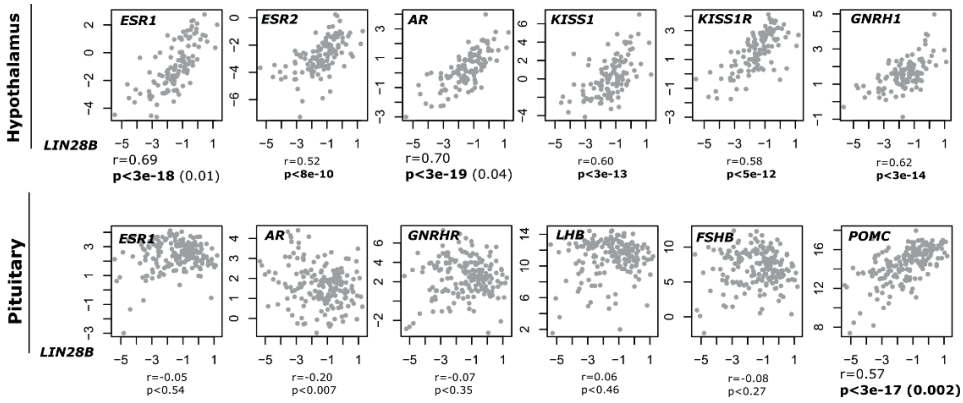


Figure 19. *LIN28B* expression correlations with selected HP axis genes from the GTEx database. In the hypothalamus, the expression of *LIN28B* correlates strongly with the expression of genes that are known to promote pubertal onset, but these correlations are not reflected in the expression of their targets *LH* and *FSH*. Conversely, pituitary *POMC* expression correlates significantly with *LIN28B* expression. P-value in parentheses reflects how likely another gene that has similar expression levels to the studied gene has as strong or stronger correlation with *LIN28B*).

5.6 *LIN28B* affects serum testosterone levels (Study III)

Collectively, the available GWAS data from humans, and the results from our zebrafish models and the GTEx database suggest that *LIN28B* may affect the expression of hormonal genes at the HP axis in an unexpected manner. Therefore, we finally wanted to explore the possibility that the gene might contribute to serum hormone levels in adults. Previously, we have shown how pubertal timing delaying allele (C) for rs7759938 from the *LIN28B* locus associates with higher *LIN28B* expression in the hypothalamus and the pituitary in adult humans. To finalize study III, we performed an association analysis to detect whether the same marker would affect serum testosterone levels in the UKBB database. Remarkably, we observed that the C allele at rs7759938 associates robustly with lower serum testosterone levels in adult humans (N=351,062, P= 9.2e-16, **Table 10**). The association was especially clear for males (*P*=2.5E-37) whereas it did not reach genome-wide significance (P<5e-8) in females (*P*=0.0067). In study III we thus provided for the first time evidence that the variants associated with higher *LIN28B* expression associate also with lower serum testosterone levels in adult humans.

Table 10. Association of rs7759938 with serum testosterone levels in UKBB. A1=effect allele, A2=other allele, EAF=effect allele frequency.

	SNP	N	Chr	Pos	A1/A2	EAF	Beta	SE	P
Males	rs7759938	176212	6	105378954	C/T	0.32	-0.044	0.003	1.5E-37
Females	rs7759938	174850	6	105378954	C/T	0.32	-0.010	0.003	0.0067
M+F Combined	rs7759938	351062	6	105378954	C/T	0.32	-0.020	0.002	9.2e-16

5.7 Discussion: insights in to the function of *LIN28B*

The previous chapters describe and summarize the key findings from the original studies I-III forming the basis of this thesis. Each of the original studies has provided novel information about the function of *LIN28B*. Although we still lack a definite answers on for example from where does the pubertal timing association come from, collectively these studies have increased our understanding of the potential molecular mechanisms by which *LIN28B* may affect the formation of complex traits. In the beginning of this thesis project, we had set three specific aims. The aims included 1) studying how *LIN28B* affects body size and metabolism in humans, 2) in which way and in which tissues do the sequence variants in the *LIN28B* region affect the gene's expression and 3) how does manipulating *lin28b* expression in zebrafish affect the development and growth of the fish. On the way, the projects evolved to encompass also studying *LIN28B's* potential contribution to hormonal signaling in vertebrates. Below I discuss whether these aims were accomplished, and what additional implications these results may have.

5.7.1 LIN28B has pleiotropic effects on body size but has little effect on serum metabolites in adult humans

The aim 1, studying how *LIN28B* affects body size and metabolism in humans, was mostly addressed in study I. Overall, the results from study I, encompassing ~26,000 Finnish individuals suggested that *LIN28B* has potential to contribute to body size and shape in more complex ways than appreciated at the time. We associated variants rs7759938 and rs314279, linked with delayed pubertal timing, with increased height, weight, hip size and WHR. Since then, studies in larger datasets have provided more evidence that sequence variants in the *LIN28B* region indeed affect human body size in complex ways (144, 145). Remarkably, the studies by others and us have also shown how homologous *LIN28* genes may affect organism size also in complex ways in other vertebrates like zebrafish and mice, strongly supporting the notion that *LIN28* genes participate into evolutionarily conserved genetic pathways that regulate development and growth in vertebrates. For example, it

appears that zebrafish with modified *lin28b* expression exhibit alterations in body size. This and the potential mechanisms behind the body size associations will be further discussed especially in chapter **5.7.3**, addressing the results from the zebrafish models.

Although *LIN28B* contribution to human body size therefore appears robust, the results from our study did not show evidence for *LIN28B* significantly affecting lipid nor T2D related metabolite levels in adults, although we detected a nominal association with rs314279 and insulin levels in males. Many of the body size associations from our study have been later replicated in larger datasets, but evidence for *LIN28B* affecting lipids and T2D related traits is still lacking despite some considerably larger studies have been conducted (19). Considering the reports that *LIN28B* expression may affect the choice of metabolic pathways in cell models, mice and apparently in zebrafish, this is slightly surprising (116, 131). Notably, however, for example the mice models addressing *Lin28* effects on glucose metabolism are based on leaky overexpression or complete knockout of *LIN28* function. Contrastingly, under normal conditions *LIN28B* expression peaks during embryogenesis, and becomes restricted to specific cellular structures in mature tissues (203). Therefore, we can speculate that any effects *LIN28B* might have on metabolism would be restricted to the early developmental period, and that other genetic and environmental factors are more likely to affect serum metabolite levels in adults.

One aspect that was briefly touched in study I, namely the existence of more than one causal sequence variant in the *LIN28B* locus has received little attention beyond our studies, and should be confirmed in larger datasets, ideally based on different ethnic groups. However, up to date most other published studies including large number of subjects have reported that the most prominent genetic effect in the *LIN28B* region is tagged by a haplotype that that contains the variant rs7759938 (**Table 2**). Although none of these studies have provided evidence for any of non-coding variants at the *LIN28B* region for being truly causal, at the same time no protein coding variants in the *LIN28B* gene (that remain extremely rare) have been shown to have phenotypic effects. The challenge to pinpoint the actual causal variant(s) in the *LIN28B* locus thus still exists. However, most likely the main genetic effect underlying the GWAS signals is based on a variant affecting *LIN28B* expression: in study II we showed how different alleles for rs7759938 correlate with differences in the amount of *LIN28B* mRNA transcript in the HP axis (discussed in more detail in the chapter **5.7.2**).

Besides offering evidence of the pleiotropic actions of *LIN28B*, study I highlights a couple more points relating to studying GWAS loci. First, it seems that despite not reaching "genome-wide-significant" *P*-values, even weaker associations may reflect true biological effects. Especially, this is likely true if there exists a solid biological hypothesis behind the association, like in the case of *LIN28B* and hip circumference (based on mice studies, *LIN28* genes appeared to lead to a general increase in body size). The contribution of such weak genetic effects on complex traits has been extensively theorized and also shown to apply in many cases (86). On the other hand, for many phenotypes in our dataset the effect size estimates showed wide confidence intervals and varied between markers and depending on sex. Whereas some of the associations we detected have been replicated, others have not. For example, after our study no association with *LIN28B* and insulin levels has come up, even the number of individuals in insulin GWAS has risen to hundreds of thousands in the largest studies (19). Altogether, this could be seen as a demonstration of the fact that when dealing with subtle genetic effects, a relatively large study like ours (encompassing ~4000-26,000 subjects depending on the phenotype) has power to detect some effects, but may still remain sensitive for statistical fluctuation or uncontrolled demographic and environmental effects. Notably, the study I also lacks formal control of potential population stratification for example by including genetic principal components (PCs) in the analysis as covariates, which in theory may lead to spurious associations. However, in general we saw little evidence for the association patterns differing markedly between different geographical regions, or collection years of the FINRISK studies (data not shown).

5.7.2 Gene expression data supports *LIN28B* as the likely causal gene in the locus and links the gene with the regulation of the HPG axis

The second aim of this thesis, assessing how the sequence variants in the *LIN28B* locus affect gene expression and whether this might have consequences on the function of the HPG axis was addressed in studies II and III. To this end, we leveraged the gene expression data from the GTEx database, containing tissue specific gene expression data from adult humans in both studies. In study II, reporting the results for SNP rs7759938 we showed how the pubertal timing associated genetic variation likely affects *LIN28B* expression mainly in the HP

axis of adult humans. Our study was the first to show a direct link between a GWAS SNP and *LIN28B* mRNA levels, suggesting that understanding *LIN28B* function at the HP axis may be critical for understanding the biology behind the GWAS associations. Importantly, although previous animal models had strongly suggested that *LIN28B* might be the causal gene in the locus (instead of the other genes that reside near the GWAS SNPs) the study was the first one to show that *LIN28B* is the only protein coding gene whose expression if affected by these SNPs. The result thus strengthens the concept of *LIN28B* being the truly causal gene in the locus. The most up to date versions of the GTEx database, including over 700 individual donors, have only reinforced the reported association between pubertal timing markers and *LIN28B* expression (169). Yet, it is worth recalling that these results from the GTEx project do not necessarily reflect the transcriptional milieu that is present around the timing of puberty in healthy population, as the database is mostly based on relatively old (median age >50), deceased individuals. However, we can speculate that these changes in *LIN28B* expression tagged by these SNPs indeed remain consistent throughout human life-course, and thus might underlie some of the genetic association results.

In study III we found out how *LIN28B* expression correlates positively with the expression of several hormonal genes in the HP axis. This was somewhat unexpected, since for example the activity of *KISS1* and *GNRH1* is known to promote sexual maturation – which is a phenotype related to lower *LIN28B* expression. The parallel observation that decreased *LIN28B* expression did not correlate with decreased gonadotrophin mRNA levels in the pituitary anyhow points towards an interesting hypothesis that might explain why *LIN28B* associated with traits like puberty and 2D4D finger length ratio (**Figure 20**). Potentially, with lower *LIN28B* expression the pituitary might become "calibrated" to require less hypothalamic stimulation to induce gonadotrophin secretion. In theory such an effect might be driven for example by *LIN28B* affecting the cellular structure and organisation of the hypothalamus/pituitary, instead of directly controlling the expression of individual hormonal genes. This would offer an intriguing mechanistic explanation for the advanced onset of puberty associated with lower *LIN28B* expression: the individuals that express less *LIN28B* would be poised to start gonadotrophin secretion with less hypothalamic stimuli, i.e. lower GnRH pulse would be required for pubertal onset. On the other hand, by reducing the hypothalamic *ESR1* and *AR* expression, the data allows also speculation that *LIN28B* might

simply modify the sensitivity of the HP-axis to negative sex steroid feedback, causing an increase in the amount of circulating sex steroids. However, these mechanisms remain speculative and require further study, and for example do not overrule the hypothesis that *LIN28B* effects on pubertal timing might be related simply to somatic maturation of the HP axis (134). Nevertheless, these hypotheses resonate well with the result that the same SNPs that increase *LIN28B* expression also decrease testosterone levels in adult humans, in both sexes (section **5.6**, Study III). The theories would go together also with the observation that the timing of puberty is similarly altered in males and females, supporting the notion that the change in the timing is driven by a central mechanism affecting both sexes (10, 210).

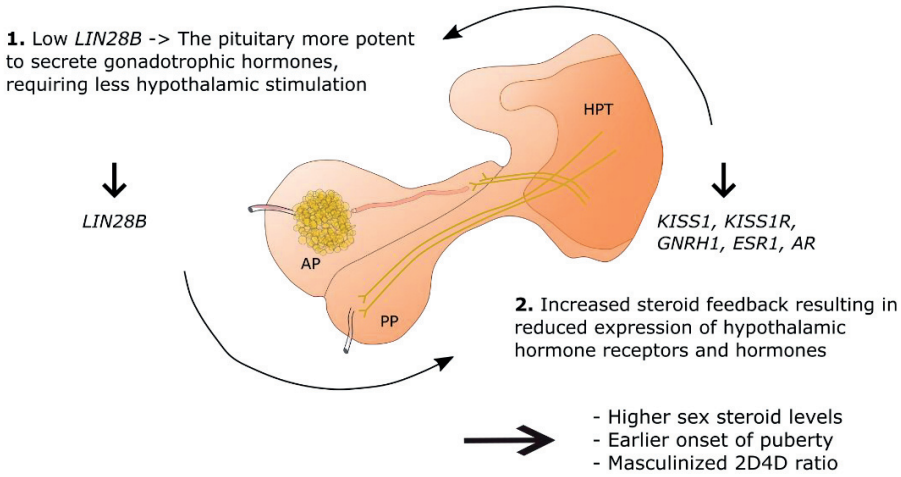


Figure 20. Schematic description of a speculative mechanism whereby *LIN28B* might affect the hormonal dynamics at the HP axis based on GTEx expression data and GWAS associations.

Another observation from the GTEx data in study III (higher *LIN28B* expression associates with higher *ESR1* expression) is also particularly interesting for several reasons. First, the observation that estrogen signalling may be altered in the HP axis raises the possibility that the effects *LIN28B* has on sex steroids are not be limited to testosterone, but may also involve estrogen and estradiol. Additionally, this reflects some of the findings from the zebrafish models. As discussed in the section **5.7.3**, estrogen signalling appeared affected

also in the *lin28b* KO fish, supporting the potential relevance of *LIN28B* to the regulation of sex steroid pathways seen in human data. Secondly, both *ESR1* and *LIN28B* have been shown to affect similar traits. Mice studies have shown how estrogen receptor alpha (ESR*α)* signalling in kisspeptin neurons regulates puberty: the mice in which *ESRα* has been knocked out in KISS neurons show advanced pubertal development (211). Moreover, *ERα* knockout in medial basal hypothalamus has been shown to lead to increased bone mineral density (BMD) in female mice. Fascinatingly, the genetic variants that lower *LIN28B* expression have been associated with both traits, supporting the idea that the *ESR1s* finding might be particularly important for understanding *LIN28B*-related biology. Nonetheless, the regulation of pubertal onset and progress by estrogen signalling is notoriously complex, including both suppressive and activating functions and multiple receptors, and hence the finding should be carefully followed up (25, 211).

Finally, based on the GTEx data an interesting observation was the strong positive correlation between *POMC* and *LIN28B* expression in the pituitary. Captivatingly, increased *POMC* activity has been linked to lower prenatal testosterone exposure, fitting together with the observation that higher *LIN28B* expression associates with lower testosterone levels (212). Lower prenatal testosterone in turn is known to be closely related to higher (feminized) 2D4D-ratio, one of the phenotypes that has been associated with increased *LIN28B* expression (213). Additionally, one of the peptides made based on *POMC mRNA* is b-endorphin, which binds opioid receptors and participates to regulation of stress response, and thus might be relevant in terms of *LIN28B* association with depression and potentially also insomnia (17). In essence, the gene expression data from the GTEx thus indicates that *LIN28B* expression might actually modify the hormonal dynamics at the HP axis in complex and previously unappreciated ways. Overall, the results from the database have provided critical insight into the possible mechanisms that might underlie *LIN28B* associations with pubertal timing and other complex traits.

5.7.3 Zebrafish models suggest *LIN28B* affects growth in an evolutionarily conserved manner, and hint at possible molecular mechanisms behind the genetic associations

Besides using several different types of human data, the original studies of this thesis have largely revolved around various zebrafish models. For aim three, to study *lin28b* function in the regulation of vertebrate growth, we generated transient *lin28b* knockdown/overexpression models in study II, and permanent *lin28b* KO models in study III. Taken together, the results from the studies support the concept that *LIN28*-genes have potential to affect vertebrate growth in a complex manner. Remarkably, it appears that the genetic pathways that involve *LIN28B* have the potential to define the patterns and limits for vertebrate growth already during embryogenesis. In study II we had three main findings: 1) Increasing *lin28b* expression transiently by synthetic mRNA injections during embryogenesis results in permanent increase in zebrafish size 2) *lin28b* may have potential to contribute to development of the zebrafish hypothalamus and 3) the mechanisms causing *LIN28B* to associate with body size and pubertal timing might be distinct. Considering also the data from mice, and the established expression pattern of *LIN28*-genes, the results from study II strengthen the concept that the effects these genes have on growth may be determined relatively early during development, when the expression of the genes is ubiquitous and strong compared to adult tissues. The result is supported by the finding in mice that fetal actions of *Lin28B* are crucial for growth (131).

Arguably, the studies in zebrafish support the concept that the genetic effects by which *LIN28B* affects body size may be distinct from the effects affecting pubertal timing. Importantly, these may be separated in time: we now think *LIN28Bs* effect on growth potential may be largely determined during the fetal stages, but the effect on pubertal timing might occur later during development. First, in study II transient overexpression of *lin28b* during embryogenesis increased zebrafish size as expected, but had rather unanticipated effects on the sexual development of the female fish. In study II, embryonic *lin28b* overexpression causing bigger body size also advanced sexual maturation, which we interpreted as a secondary consequence of the bigger body size, and not as a direct *lin28b* effect. Contrastingly, in study III assessing the consequences of permanent *lin28b* knockout

we observed both reduced fish growth and – using growth velocity as a proxy – also the anticipated effect (advancement) on the timing of sexual maturation.

To offer insight into the genetic mechanisms regulated by *lin28b* during early development (potentially affecting fish growth), in study III we performed a global RNA-seq analysis on 1dpf and 7dpf fish. Remarkably, it appeared that mitochondrial metabolism was enhanced in the *lin28b* KOs during these early stages, which in theory might be mechanistically related to the smaller size of the KO fish. We concluded that, like speculated in the case of *Lin28A* and mice, the effect *LIN28* genes have on growth might be related to the regulation of overall number of cells across tissues. This in turn, might be achieved by *LIN28B's* contribution to cellular metabolism: suppression of mitochondrial OXPHOS at the expense of glycolytic pathways, (that is known to be the preferred metabolic choice of dividing cells) (116, 208). Overall these results from zebrafish thus appear to strengthen the concept that *LIN28*-genes have potential to affect cellular metabolism during early development. Unfortunately, we lacked models for addressing whether these effects are indeed restricted to early development in zebrafish.

Besides addressing phenotypes linked with cellular growth, many of the phenotypes associated with *LIN28B* are subject to hormonal regulation and the function of the HPG axis. Therefore, with our models we additionally wanted to specifically assess whether tinkering with *lin28b* expression might have consequences on the function of the hypothalamus and the pituitary in zebrafish. The results based on our models are partly conflicting: in study II we observed that knocking down *lin28b* might delay the development of the hypothalamus (reflected by suppression of *kiss2* expression), but in *lin28b* KO fish the development of the axis appeared relatively normal, or rather more advanced. Based on these models it thus appears that *lin28b* may not be essential for development of the HP axis, but may indeed have potential to affect its function. Partly, the conflicting results between the two models might be related to the fundamental differences that exist between them: compared to gene knockdown by MO used in study II, which is notoriously prone to causing off-target effects, gene KO strategy used in study III is contrastingly notoriously well compensated in zebrafish (179, 184).

Suggesting that *lin28b* function might be relevant for the HP axis function, we however observed that hormonal signaling might be altered also in *lin28b* KO fish. Already at 7dpf, when *lin28b* expression in largely restricted to the hypothalamus and the pituitary, GO pathway analyses suggested that the KO fish show upregulation of genes that respond to estrogen signaling. Similarly, at 130dpf (around the timing of sexual maturation) the main estrogen receptor (*esr1*) appeared downregulated in the KO brain. The findings thus support the notion that sex steroid signaling might be upregulated in the KO fish. They also resonate with the data from the GTEx, which links sequence variants with lower *LIN28B* expression to downregulation of *ESR1* in the hypothalamus in humans. Therefore, simultaneously considering the testosterone finding discussed below in section **5.7.4**, it currently appears likely that the contribution of *LIN28B* to hormonal regulation may be crucial to explain why the gene has been associated with many traits such as pubertal timing, 2D4D finger length ratios and BMD.

Overall, the results from *lin28b* zebrafish models have aided in creating a picture about the significance of *LIN28B* for vertebrate development, and have partly revealed potential molecular mechanisms whereby the gene might affect several phenotypes. In summary, the results indicate that *LIN28B* related genetic mechanisms regulate the timing and tempo of vertebrate growth in an evolutionary conserved manner. Importantly, based on the transient overexpression models, it seems that *LIN28B* related genetic pathways represent a set of genetic mechanisms, which – if perturbed during embryogenesis due to any reason – have the potential to impact the subsequent developmental framework over an organism's life course. Moreover, based on these zebrafish models, is appears possible that the pleiotropy of *LIN28B* can be partly explained by different actions of the gene at different time points in development. It appears likely that *lin28b's* effects on cellular metabolism during the fetal stages may directly affect the regulation of vertebrate body size, whereas the potential effects the gene has on hormonal signaling later in life might be more relevant for phenotypes like puberty and sexual maturation.

5.7.4 Could the testosterone association help to explain why *LIN28B* associates with so many phenotypes?

Over the course of this thesis, it has become increasingly apparent that *LIN28B* and its homologues in other species may have potential to contribute to hormonal dynamics in vertebrates. Remarkably, the finding that the SNPs linked with higher *LIN28B* expression associate also with lower circulating testosterone levels in humans might be crucial in terms of answering the question that essentially motivated this project: how does *LIN28B* affect pubertal timing (and many other traits that have been since associated with the gene). Over the past years, many data sources have offered indirect evidence that *LIN28B* might affect the hormonal milieu for an individual. First, many of the traits that have been associated with *LIN28B* are directly regulated by hormones. For example pubertal timing depends on the acquisition of hypothalamic GnRH pulse that translates to gonadotrophin secretion, triggering sex steroid activity that is ultimately responsible for the physiological changes associated with puberty. Already before the testosterone association result, based on the GTEx data, we had speculated that the positive correlation with *LIN28B* and hypothalamic gene expression (discussed in results section **5.5**) might reflect changes in circulating sex steroid levels, a hypothesis that has now gotten some further support. Notably, although the effect was stronger in males, we saw evidence of *LIN28B* affecting testosterone levels in both sexes. Combining this observation with the result from study II (the SNPs at the *LIN28B* region affect *LIN28B* expression primarily in the hypothalamus and pituitary and not in the testis), we have grounds to speculate that the gene's effect on testosterone levels might actually be set at the HP axis. Like speculated in the chapter **5.7.2**, this might result from changes in gonadotropin secretion or HP axis response to negative steroid feedback. This again raises the possibility that the effect is not limited to male hormone testosterone: *LIN28B* might also affect estrogen levels, although this remains to be studied.

Besides puberty, many other traits linked with *LIN28B* have shown links to sex steroids. The gene has been associated with traits like 2D4D finger length ratio and bone mineral density, which have both been shown to be regulated by testosterone (213, 214). As in the case of puberty, the direction of the effect is biologically sensible i.e. SNPs associating with lower *LIN28B* expression (and thus higher testosterone levels) associate also with masculinized 2D4D ratio and increased BMD, traits which have been directly shown to be affected by the

hormone (214, 215). Although we lack any direct causal evidence that *LIN28B's* effects on testosterone indeed underlie these associations, the results presented in this thesis thus strongly suggest that the gene may act as a general regulator of the HP-axis function controlling hormone levels in humans, and further studies on how *LIN28B* contributes to hormonal regulation in vertebrates are therefore warranted. Overall, considering the vast potential of the HPG-axis to regulate human development, it however appears feasible that *LIN28B*'s effects on some of the phenotypes might indeed be mediated through effects on this pathway.

6 Concluding remarks and future prospects

The previous chapters have introduced the field of genetics of pubertal timing, summarized current knowledge about *LIN28B* and presented the results from the original studies serving as the foundation for this thesis project. During this thesis work, it has become evident that *LIN28B* is not only a pubertal timing gene, but rather a gene that contributes to the formation of multiple, seemingly unconnected traits across vertebrate species. With the studies presented in this thesis, we have shown that besides affecting pubertal timing *LIN28B* has pleiotropic effects on human body size, highlighted the potential significance of the gene's actions for the function of the hypothalamic-pituitary axis, and contributed to the establishment of *LIN28B* genes as evolutionarily conserved regulators of vertebrate growth.

In general, the results from our studies have been largely consistent with other studies assessing *LIN28B* function. For example the effects of *LIN28B* on body size appear particularly robust and replicable across vertebrates. Large GWAS studies now suggest that the genetic variants that decrease *LIN28B* expression lead to a proportional decrease in body size – an effect that we observed and report both in the human data and in our zebrafish models. In addition, similarly to the data from *Lin28B* mice models, our zebrafish data indicates that this decrease in size is highly related to the actions of the gene during early development, namely embryogenesis and fetal development. These in turn appear to be closely related to changes in cellular metabolism. Importantly, our studies have also provided completely novel insight into *LIN28B* function. We have showed how the pubertal timing associated SNPs affect *LIN28B* expression primarily at the HP axis in adult humans, and suggested that in the same tissues *LIN28B* may contribute to gene expression dynamics also after the fetal stages. In addition, our studies show how *LIN28B* may directly affect serum testosterone levels in adult humans, which might be relevant in terms of understanding how the gene contributes to various phenotypes including pubertal timing. Altogether, these discoveries have opened up novel directions to explore the molecular mechanisms that regulate vertebrate growth later in life. For example, as *LIN28Bs* effects on growth appear robust and replicable in all the studied species, it would be tempting to study next whether the *LIN28B*-related pathways might contribute also to variation in growth tempo and amplitude between species.

In addition to providing insight into *LIN28B* function, in the larger context this thesis has served as an example of studying the GWAS loci in general. Understanding the molecular background of the huge number of genetic associations that are currently available presents a big challenge. Even the question of whether it makes sense to try to make sense of all these associations with the current methods is valid, especially considering the relatively small proportion of trait variance an individual locus usually explains. Nevertheless, targeted studies may be warranted in particular if the genetic effect on the phenotype is large in the GWAS terms, and in case these have potential to reveal novel biological pathways behind the formation of complex traits – as was the case with *LIN28B* and pubertal timing in the beginning of this project.

In the wake of this project, there are some key considerations that might be generalizable to other projects aiming study GWAS loci in detail. First, studying gene expression may be a logical starting point for studying GWAS loci, as this may provide vital clues about the direction of the genetic effect and the causal tissues. Secondly, identification of the causal SNP in the region obviously may provide additional information about the molecular mechanisms, but is not absolutely necessary for further studies, particularly in case this would require considerable effort. Thirdly, understanding the "phenome-wide" consequences of a SNP on the other hand may offer vital cues about the function of a gene. In case one does not want to rely solely on published studies and does not have direct access to private cohort data, large scale biobanking initiatives like UK biobank now offer publicly available genetic association data on thousands of traits (19). Finally, it appears that although in many cases the original associations may stem from relatively small differences in gene expression, "Mendelian" knockout/knock-in animal models may still be very useful in studying the molecular background of these associations. Despite the genetic lesion introduced in these models is usually more severe than the original variant causing the association (and in theory might lead to unexpected phenotypes), in many instances they still replicate the core features of the original association, like shown in this project. In fact, continuing to rely on the Mendelian models when studying the GWAS loci in animal models may be considered somewhat essential. The GWAS hits are based on studying large databases with thousands of human samples, and thus it would not be practical to try to mimic the observed effect size in animal models: to detect anything under such models

would likely require an impracticable/unethical amount of experiments and use of animals. Moreover, when considering which model organism to choose, it appears that it does not necessarily have to be closely related to humans, especially if there is reason to suspect that the association reflects an evolutionarily conserved genetic mechanism, like in the case of *LIN28B*. In general terms, however, perhaps the primary lesson that studying *LIN28B* has thought us is that there is a lot to be gained if it is possible to gather data from multiple sources, instead of just concentrating on studying a GWAS locus solely from one angle. Mostly, this has been made possible by public data sharing in the genetic community. In our project we have collected data from large population cohorts, gene expression databases and sequencing databases and combined this information with results from various different animal models including our own zebrafish models - simply to understand why certain genetic variants out of the millions of variants in our genome are linked to differences in pubertal timing in humans. By taking such a broad view, it is additionally easier to put the individual findings from this thesis into a bigger perspective: essentially this thesis project has not been dealing just with *LIN28B* expression levels and puberty in humans. Instead, we have been working to uncover mechanisms that regulate development, growth and disease and that may affect phenotypes, metabolism and ontogeny in a wide range of metazoans. Importantly, this thesis project can be seen also as a general example of the potential ways the GWAS findings can be followed up.

Besides offering general guidelines into studying GWAS loci, the results from our project touch another timely topic in genetics. We are entering an era where human gene editing and "designer babies" have become reality. The experiences gained during this project underscore two central paradigms that all responsible researchers involved in genetic engineering should acknowledge. First, it is quite evident that the knowledge about the function of individual genes is still largely limited. For example, we started with *LIN28B* as a pubertal timing gene, but throughout this project, new associations with *LIN28B* and several human traits have constantly emerged. Secondly, following these associations, it is evident that *LIN28B,* like many genes, has pleiotropic effects on human phenotypes. It is thus naïve to think that by editing a single gene one would only affect a single trait. In the case of *LIN28B*, would one then prefer a version of the gene that increases height, intraocular

pressure and cancer risk, or the one that increases bone mineral density and risk for insomnia and depression?

In the end, a logical, though slightly awkward question remains. This thesis was set to explore the molecular mechanisms behind the GWAS results - so do we now really know how *LIN28B* affects puberty? And why might *LIN28B*, a stem cell factor controlling cell faith be related to all the other phenotypes that have been associated with the gene? Although we now certainly know more about *LIN28B* function than we did few years ago, many questions remain. For example, although *LIN28B* expression levels seem to correlate with vertebrate body size, potentially through effects on cellular metabolism and cell differentiation, the question - how does this actually happen – is still valid. On the other hand, we also recognize that *LIN28B* function at the HP axis and the gene's contribution to regulation of hormone levels may be important for many of the phenotypes associated with the gene. Yet, whether this happens for example due to *LIN28B* affecting the somatic maturation and development of the HP axis, because it directly affects the expression of specific genes, or because it causes structural changes in the organization of these tissues remain unsolved: in theory all the above mechanisms could affect pubertal timing. Likewise, although we currently have evidence that *LIN28B* contributes to serum testosterone levels, which in theory might explain some of the GWAS findings, we still lack causal evidence for this as the mechanism behind the associations. Therefore, to protect oneself from the potential depressive consequences of being homozygous for the T allele of rs7759938 at the *LIN28B* region, I will leave these questions aside for a while. Instead, I will next concentrate on enjoying the fact that this thesis is now complete, and do my best to try to thank everybody who has been involved in this journey.

7. Acknowledgements

The research presented in this thesis was carried out at the Institute for Molecular Medicine Finland, (FIMM) and at the Department of Anatomy of the University of Helsinki. I wish to acknowledge Olli Kallioniemi, Jaakko Kaprio and Mark Daly as directors of FIMM for providing excellent research facilities, infrastructure and scientifically stimulating working environment.

This thesis has received funding from the Academy of Finland for projects "Genetic Regulation of the Timing of Puberty", "Elucidating Early Developmental Effects of Loci Associated with Pubertal Growth and Maturation" and the Center of Excellence in Complex Disease Genetics (CECDG) here at FIMM. Moreover, I have been personally funded by the University of Helsinki Doctoral School in Health Sciences, Doctoral Programme in Biomedicine (DPBM) and the Finnish Cultural Foundation. I would also like to acknowledge the support from Ylikiimingin Nuijamiehet for extracurricular activities. I am truly grateful to all these sources for their financial support and their contribution to science.

I want to express my deepest gratitude to my wonderful supervisor Elisabeth Widén who has made the entire project possible and supported me and my development throughout the time I've spent here. I have been lucky to be teamed up with a person who understands both science and life and all the complexities within. Showing your appreciation for things like family and sports has meant a lot to me. Thank you for sensitively guiding my thinking and gently but patiently pushing me forward during all these years. I am especially grateful for you giving me the independence and the chance to participate into several interesting projects also outside my thesis, to try, fail and sometimes succeed, yet always feeling that you support my choices and trust me and my ability to make things happen.

The pre-examiners of my thesis, Reetta Hinttala and Craig Primmer: thank you for accepting the responsibility and taking the time to review my thesis. I'm truly happy that the doctoral committee at the faculty of Medicine accepted you as my pre-examiners. You both provided valuable comments and expertise to improve the content, format and presentation of this thesis. I also wish to thank Prof. Alexandre Raymond for accepting the role of serving as the opponent of my defence. I'd also like to thank Samuli Ripatti for once again accepting the role of Kustos in a defence. Your contribution to make these theses happen has not gone unnoticed. Especially, and separately I want to thank you for your support to this study as the director of the CECDG.

This thesis includes relatively many instances where talking about research I use the word "we" instead of "I". I hope my co-authors appreciate why. None of this could have happened without you: Aki Havunlinna, Johannes Kettunen, Ida Surakka, Taru Tukiainen, Jana Pennonen, Nella Junna, Leevi Lehtonen, Riitta Luoto, Veikko Salomaa, Elisabeth, Yu-Chia Chen, Pertti Panula have each made

contributions for this thesis either by providing data, samples, statistical expertize, help in analysis, mentoring, resources or assistance in lab work. I would like to underscore the importance of the "Zebrafish people": the invaluable teaching, mentoring and assistance by magical Yu-Chia in designing and performing the lab experiments, and Pertti providing space, expertise and resources for this project at the Department of Anatomy and at the Zebrafish Unit. I'd also like to highlight how important my former and present colleagues have been to make things happen: without Jana, Nella and Leevi I might lack some of the most important results presented here, and instead would spend my time either struggling to get the fish to co-operate or trying to understand how does BOLT work. Similarly, the help from fabulous Taru in terms of analyzing and teaching me about the GTEx data has been crucial for this thesis. Your ambition and skills have helped me to improve my papers and drive me forward as a scientist. I am privileged to be able to continue working with you! Finally, I want to thank my lovely former colleague Diana Cousminer for paving the way for me in Elisabeth's group and helping me to navigate my way around as a noob in the field of complex disease genetics and at FIMM. It would be great if you one day ended up back to Finland! Moreover, I want to especially mention and thank Cristian Benner, Matti Pirinen, Henri Puttonen, Maria Sundvik and Lea Urpa for their kind assistance in the projects included in this thesis.

Besides the thesis project, during these years I have had the pleasure to work in several other interesting projects from which want to mention some key people here that have left a mark on me. I'd first like to thank my key collaborators from the successful arrhythmia projects: I had the honour and pleasure to work with Heikki Swan over several years, together with Hugues Abriel, Lia Crotti, Aurora Djupsjöbacka, Kimmo Kontula, Susanna Saarinen, Peter Schwartz, Annukka Tuiskula and Matti Viitasalo: thank you for introducing me to the field and I truly cherish what we accomplished! Similarly, a huge thanks goes to Nina Linder, Johan Lundin, Reija Randén-Brady and Tiina Vesterinen for all your work and helping me to understand *LIN28B* from cancer perspective. I am also very proud of the "Soft Skills Team" we founded here at FIMM and would like to thank you Tiia, Arjan, Meri, Elina, Pu and Alyce for all the great times we have had planning and running those sessions!

I'm obviously greatly indebted to the Technology Centre personnel and laboratory staff here at FIMM and at the Zebrafish Unit, who have greatly assisted me in my experiments, teaching me and in many instances providing ah so great data or fresh fish: Kati Donner, Sirkka Ekström, Pekka Ellonen, Sari Hannula, Tiina Hannunen, Eija Hämäläinen, Matti Kankainen, Suvi Kyttänen, Päivi Lahermo, Jenni Lahtela, Maija Lepistö, Pirkko Mattila, Aino Palva, Jouko Siro, Anu Suoranta, Minna Suvela, and Anu Yliperttula; Reetta Huhtala, Henri Koivula, Suski Norrbacka, Riikka Pesonen and Katarina Ven. I also want to thank the coordinators, administrative staff and IT personnel at FIMM who make things run: Olle Hansson, Sanni Hyppönen, Mari Kaunisto, Sari Kivikko, Eiri Matikainen, Timo Miettinen, Teemu

Perheetupa, Gretchen Repasky, Susanna Rosas, Huei-yi Shen, Birgit Simell, Ulla Tuomainen, and Emilia Vanamo.

Importantly, I wish to acknowledge all the spectacular junior and senior people in the FIMM Genomics Unit that have not been mentioned yet but with whom I have shared an office, exercise, coffee, science and many great moments, and who have made it enjoyable to come to office every day: thank you Aarno (who deserves a special mention also for kindly sharing his group's lab resources), Alfredo, Anni, Annu, Ansku, Antti, Ari, Clara, Elina, Elmo, Emmi, Hande, Hanna, Hannele, Heidi, Himanshu, Iiro, Jake, Jarmo, Javier, Joel, Johanna, Juha, Jukka, Julia, Kalle, Kimmo, Laxman, Liisa, Leif, Luca Maija, Mari N, Mari R, Marine, Marita, Marjis, Marjo, Mervi, Mikko, Mitja, Nina, Olli, Om, Paavo, Perttu, Peter, Pietari, Pietro, Priit, Pyry, Risto, Rubina, Saana, Sanni, Sini, Susanna, Tea, Timo, Tuomo, Vincent, Virpi, Vishal and Will, Yu and everybody who I forgot. On top of these guys I've had a chance to get to know many other fantastic people here at FIMM and at the University of Helsinki: some I've met in the locker room, some in the corridors, some over courses, some whilst doing lab work either at FIMM or at the Department of Anatomy. No names given to fit these acknowledgements on tree pages, but I hope you recognize yourselves!

In addition to spending my days thinking about genes, it has been great to maintain some contact with the outside world. Juho, knowing me through all these years, you might still have more to do with me graduating than you might first think. Your friendship is precious. Janne R, Janne U, Juha and Ville, it has been invigorating to have people who share similar passions around me, and with whom I've enjoyed some of the most memorable moments in my life, ranging from going out for a simple run to crossing one's physical limits.

Balancing family life, demanding work and serious hobbies requires that you have good people around you. I want to specifically thank my sister Katri, cousin Tiitu, sister-in-law Elli and especially my parents-in-law Jussi and Maire for all their support for our family during my thesis work. Particularly, I want to thank my mum Maire and dad Pekka for all their love, care and encouragement. I also hope you appreciate that the things I have learned by listening you and feeling, sensing and observing my surroundings as a kid have given me a biology/genetics education that cannot be taught in any university.

Finally, I want to thank my own family Outi, Otso, Topias and Emma for everything. You would deserve you own chapter in this book telling how great you are and what you mean to me. You're the best. My kids, I am so happy and proud of each of you. Outi, sometimes people have asked why I'm so happy and how do I manage everything I do. I've told that the secret is to have a great wife. I love you all so much.

8. References

1. Parent A-S, Bourguignon J-P, Juul A, Teilmann G, Skakkebaek NE, Toppari J. The Timing of Normal Puberty and the Age Limits of Sexual Precocity: Variations around the World, Secular Trends, and Changes after Migration. Endocr Rev. 2003;24(5):668-93.
2. Abreu AP, Kaiser UB. Pubertal development and regulation. The lancet Diabetes & endocrinology. 2016;4(3):254-64.
3. Palmert MR, Boepple PA. Variation in the Timing of Puberty: Clinical Spectrum and Genetic Investigation. The Journal of Clinical Endocrinology & Metabolism. 2001;86(6):2364-8.
4. Day FR, Elks CE, Murray A, Ong KK, Perry JRB. Puberty timing associated with diabetes, cardiovascular disease and also diverse health outcomes in men and women: the UK Biobank study. Sci Rep. 2015;5:11208.
5. Widén E, Silventoinen K, Sovio U, Ripatti S, Cousminer DL, Hartikainen A-L, Laitinen J, Pouta A, Kaprio J, Järvelin M-R, Peltonen L, Palotie A. Pubertal Timing and Growth Influences Cardiometabolic Risk Factors in Adult Males and Females. Diabetes Care. 2012;35(4):850-6.
6. Silventoinen K, Haukka J, Dunkel L, Tynelius P, Rasmussen F. Genetics of Pubertal Timing and Its Associations With Relative Weight in Childhood and Adult Height: The Swedish Young Male Twins Study. Pediatrics. 2008;121(4):e885-e91.
7. Palmert MR, Hirschhorn JN. Genetic approaches to stature, pubertal timing, and other complex traits. Mol Genet Metab. 2003;80(1):1-10.
8. Gajdos ZK, Hirschhorn JN, Palmert MR. What controls the timing of puberty? An update on progress from genetic investigation. Curr Opin Endocrinol Diabetes Obes 2009;16(1):16-24.
9. Widén E, Ripatti S, Cousminer DL, Surakka I, Lappalainen T, Järvelin M-R, Eriksson JG, Raitakari O, Salomaa V, Sovio U, Hartikainen A-L, Pouta A, McCarthy MI, Osmond C, Kajantie E, Lehtimäki T, Viikari J, Kähönen M, Tyler-Smith C, Freimer N, Hirschhorn JN, Peltonen L, Palotie A. Distinct Variants at LIN28B Influence Growth in Height from Birth to Adulthood. The American Journal of Human Genetics. 2010;86(5):773-82.
10. Day FR, Thompson DJ, Helgason H, Chasman DI, Finucane H, Sulem P, Ruth KS, Whalen S, Sarkar AK, Albrecht E, Altmaier E, Amini M, Barbieri CM, Boutin T, Campbell A, Demerath E, Giri A, He C, Hottenga JJ, Karlsson R, Kolcic I, Loh P-R, Lunetta KL, Mangino M, Marco B, McMahon G, Medland SE, Nolte IM, Noordam R, Nutile T, Paternoster L, Perjakova N, Porcu E, Rose LM, Schraut KE, Segre AV, Smith AV, Stolk L, Teumer A, Andrulis IL, Bandinelli S, Beckmann MW, Benitez J, Bergmann S, Bochud M, Boerwinkle E, Bojesen SE, Bolla MK, Brand JS, Brauch H, Brenner H, Broer L, Bruning T, Buring JE, Campbell H, Catamo E, Chanock S, Chenevix-Trench G, Corre T, Couch FJ, Cousminer DL, Cox A, Crisponi L, Czene K, Davey Smith G, de Geus EJCN, de Mutsert R, De Vivo I, Dennis J, Devilee P, dos-Santos-Silva I, Dunning AM, Eriksson JG, Fasching PA, Fernandez-Rhodes L, Ferrucci L, Flesch-Janys D, Franke L, Gabrielson M, Gandin I, Giles GG, Grallert H, Gudbjartsson DF, Guenel P, Hall P, Hallberg E, Hamann U, Harris TB, Hartman CA, Heiss G, Hooning MJ, Hopper JL, Hu F, Hunter DJ, Ikram MA, Im HK, Jarvelin M-R, Joshi PK, Karasik D, Kellis M, Kutalik Z, LaChance G, Lambrechts D, Langenberg C, Launer LJ, Laven JSE, Lenarduzzi S, Li J, Lind PA, Lindstrom S, Liu Y, Luan Ja, Magi R, Mannermaa A, Mbarek H, McCarthy MI, Meisinger C, Meitinger T, Menni C, Metspalu A, Michailidou K, Milani L, Milne RL, Montgomery GW, Mulligan AM, Nalls MA, Navarro P, Nevanlinna H, Nyholt DR, Oldehinkel AJ, O'Mara TA, Padmanabhan S, Palotie A, Pedersen N, Peters A, Peto J, Pharoah PDP, Pouta A, Radice P, Rahman I, Ring SM, Robino A, Rosendaal FR, Rudan I, Rueedi R, Ruggiero D, Sala CF, Schmidt MK, Scott RA, Shah M, Sorice R, Southey MC, Sovio U, Stampfer M, Steri M, Strauch K, Tanaka T, Tikkanen E, Timpson NJ, Traglia M, Truong T, Tyrer JP, Uitterlinden AG, Edwards DRV, Vitart V, Volker U, Vollenweider P, Wang Q, Widen E, van Dijk KW, Willemsen G, Winqvist R, Wolffenbuttel BHR, Zhao JH, Zoledziewska M, Zygmunt M, Alizadeh BZ, Boomsma DI, Ciullo M, Cucca F, Esko T, Franceschini N, Gieger C, Gudnason V, Hayward C, Kraft P, Lawlor DA, Magnusson PKE, Martin NG, Mook-Kanamori DO, Nohr EA, Polasek O, Porteous D, Price AL, Ridker PM, Snieder H, Spector TD, Stockl D, Toniolo D, Ulivi S, Visser JA, Volzke H, Wareham NJ, Wilson JF, The LifeLines Cohort S, The InterAct C, kConFab AI, Endometrial Cancer Association C, Ovarian Cancer Association C, consortium P, Spurdle AB, Thorsteindottir U, Pollard KS, Easton DF, Tung JY, Chang-Claude J, Hinds D, Murray A, Murabito JM, Stefansson K, Ong KK, Perry JRB. Genomic analyses identify hundreds of variants associated with age at menarche and support a role for puberty timing in cancer risk. Nat Genet. 2017;advance online publication.
11. Ambros V, Horvitz H. Heterochronic mutants of the nematode Caenorhabditis elegans. Science. 1984;226(4673):409-16.
12. Elks CE, Perry JRB, Sulem P, Chasman DI, Franceschini N, He C, Lunetta KL, Visser JA, Byrne EM, Cousminer DL, Gudbjartsson DF, Esko T, Feenstra B, Hottenga J-J, Koller DL, Kutalik Z, Lin P, Mangino M, Marongiu M, McArdle PF, Smith AV, Stolk L, van Wingerden SH, Zhao JH, Albrecht E, Corre T, Ingelsson E, Hayward C, Magnusson PKE, Smith EN, Ulivi S, Warrington NM, Zgaga L, Alavere H, Amin N, Aspelund T, Bandinelli S, Barroso I, Berenson GS, Bergmann S, Blackburn H, Boerwinkle E, Buring JE, Busonero F, Campbell H, Chanock SJ, Chen W, Cornelis MC, Couper D, Coviello AD, d'Adamo P, de Faire U, de Geus EJC, Deloukas P, Doring A, Smith GD, Easton

DF, Eiriksdottir G, Emilsson V, Eriksson J, Ferrucci L, Folsom AR, Foroud T, Garcia M, Gasparini P, Geller F, Gieger C, Consortium TG, Gudnason V, Hall P, Hankinson SE, Ferreli L, Heath AC, Hernandez DG, Hofman A, Hu FB, Illig T, Jarvelin M-R, Johnson AD, Karasik D, Khaw K-T, Kiel DP, Kilpelainen TO, Kolcic I, Kraft P, Launer LJ, Laven JSE, Li S, Liu J, Levy D, Martin NG, McArdle WL, Melbye M, Mooser V, Murray JC, Murray SS, Nalls MA, Navarro P, Nelis M, Ness AR, Northstone K, Oostra BA, Peacock M, Palmer LJ, Palotie A, Pare G, Parker AN, Pedersen NL, Peltonen L, Pennell CE, Pharoah P, Polasek O, Plump AS, Pouta A, Porcu E, Rafnar T, Rice JP, Ring SM, Rivadeneira F, Rudan I, Sala C, Salomaa V, Sanna S, Schlessinger D, Schork NJ, Scuteri A, Segre AV, Shuldiner AR, Soranzo N, Sovio U, Srinivasan SR, Strachan DP, Tammesoo M-L, Tikkanen E, Toniolo D, Tsui K, Tryggvadottir L, Tyrer J, Uda M, van Dam RM, van Meurs JBJ, Vollenweider P, Waeber G, Wareham NJ, Waterworth DM, Weedon MN, Wichmann HE, Willemsen G, Wilson JF, Wright AF, Young L, Zhai G, Zhuang WV, Bierut LJ, Boomsma DI, Boyd HA, Crisponi L, Demerath EW, van Duijn CM, Econs MJ, Harris TB, Hunter DJ, Loos RJF, Metspalu A, Montgomery GW, Ridker PM, Spector TD, Streeten EA, Stefansson K, Thorsteinsdottir U, Uitterlinden AG, Widen E, Murabito JM, Ong KK, Murray A. Thirty new loci for age at menarche identified by a meta-analysis of genome-wide association studies. Nat Genet. 2010;42(12):1077-85.

13. Perry JRB, Day F, Elks CE, Sulem P, Thompson DJ, Ferreira T, He C, Chasman DI, Esko T, Thorleifsson G, Albrecht E, Ang WQ, Corre T, Cousminer DL, Feenstra B, Franceschini N, Ganna A, Johnson AD, Kjellqvist S, Lunetta KL, McMahon G, Nolte IM, Paternoster L, Porcu E, Smith AV, Stolk L, Teumer A, Tsernikova N, Tikkanen E, Ulivi S, Wagner EK, Amin N, Bierut LJ, Byrne EM, Hottenga J-J, Koller DL, Mangino M, Pers TH, Yerges-Armstrong LM, Hua Zhao J, Andrulis IL, Anton-Culver H, Atsma F, Bandinelli S, Beckmann MW, Benitez J, Blomqvist C, Bojesen SE, Bolla MK, Bonanni B, Brauch H, Brenner H, Buring JE, Chang-Claude J, Chanock S, Chen J, Chenevix-Trench G, Collee JM, Couch FJ, Couper D, Coviello AD, Cox A, Czene K, D/'adamo AP, Davey Smith G, De Vivo I, Demerath EW, Dennis J, Devilee P, Dieffenbach AK, Dunning AM, Eiriksdottir G, Eriksson JG, Fasching PA, Ferrucci L, Flesch-Janys D, Flyger H, Foroud T, Franke L, Garcia ME, Garcia-Closas M, Geller F, de Geus EEJ, Giles GG, Gudbjartsson DF, Gudnason V, Guenel P, Guo S, Hall P, Hamann U, Haring R, Hartman CA, Heath AC, Hofman A, Hooning MJ, Hopper JL, Hu FB, Hunter DJ, Karasik D, Kiel DP, Knight JA, Kosma V-M, Kutalik Z, Lai S, Lambrechts D, Lindblom A, Magi R, Magnusson PK, Mannermaa A, Martin NG, Masson G, McArdle PF, McArdle WL, Melbye M, Michailidou K, Mihailov E, Milani L, Milne RL, Nevanlinna H, Neven P, Nohr EA, Oldehinkel AJ, Oostra BA, Palotie A, Peacock M, Pedersen NL, Peterlongo P, Peto J, Pharoah PDP, Postma DS, Pouta A, Pylkas K, Radice P, Ring S, Rivadeneira F, Robino A, Rose LM, Rudolph A, Salomaa V, Sanna S, Schlessinger D, Schmidt MK, Southey MC, Sovio U, Stampfer MJ, Stockl D, Storniolo AM, Timpson NJ, Tyrer J, Visser JA, Vollenweider P, Volzke H, Waeber G, Waldenberger M, Wallaschofski H, Wang Q, Willemsen G, Winqvist R, Wolffenbuttel BHR, Wright MJ, Australian Ovarian Cancer S, The GN, kConFab, The LifeLines Cohort S, The InterAct C, Early Growth Genetics C, Boomsma DI, Econs MJ, Khaw K-T, Loos RJF, McCarthy MI, Montgomery GW, Rice JP, Streeten EA, Thorsteinsdottir U, van Duijn CM, Alizadeh BZ, Bergmann S, Boerwinkle E, Boyd HA, Crisponi L, Gasparini P, Gieger C, Harris TB, Ingelsson E, Jarvelin M-R, Kraft P, Lawlor D, Metspalu A, Pennell CE, Ridker PM, Snieder H, Sorensen TIA, Spector TD, Strachan DP, Uitterlinden AG, Wareham NJ, Widen E, Zygmunt M, Murray A, Easton DF, Stefansson K, Murabito JM, Ong KK. Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. Nature. 2014;514(7520):92-7.

14. Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F, Willer CJ, Jackson AU, Vedantam S, Raychaudhuri S, Ferreira T, Wood AR, Weyant RJ, Segrè AV, Speliotes EK, Wheeler E, Soranzo N, Park J-H, Yang J, Gudbjartsson D, Heard-Costa NL, Randall JC, Qi L, Vernon Smith A, Mägi R, Pastinen T, Liang L, Heid IM, Luan Ja, Thorleifsson G, Winkler TW, Goddard ME, Sin Lo K, Palmer C, Workalemahu T, Aulchenko YS, Johansson Å, Carola Zillikens M, Feitosa MF, Esko T, Johnson T, Ketkar S, Kraft P, Mangino M, Prokopenko I, Absher D, Albrecht E, Ernst F, Glazer NL, Hayward C, Hottenga J-J, Jacobs KB, Knowles JW, Kutalik Z, Monda KL, Polasek O, Preuss M, Rayner NW, Robertson NR, Steinthorsdottir V, Tyrer JP, Voight BF, Wiklund F, Xu J, Hua Zhao J, Nyholt DR, Pellikka N, Perola M, Perry JRB, Surakka I, Tammesoo M-L, Altmaier EL, Amin N, Aspelund T, Bhangale T, Boucher G, Chasman DI, Chen C, Coin L, Cooper MN, Dixon AL, Gibson Q, Grundberg E, Hao K, Juhani Junttila M, Kaplan LM, Kettunen J, König IR, Kwan T, Lawrence RW, Levinson DF, Lorentzon M, McKnight B, Morris AP, Müller M, Suh Ngwa J, Purcell S, Rafelt S, Salem RM, Salvi E, Sanna S, Shi J, Sovio U, Thompson JR, Turchin MC, Vandenput L, Verlaan DJ, Vitart V, White CC, Ziegler A, Almgren P, Balmforth AJ, Campbell H, Citterio L, De Grandi A, Dominiczak A, Duan J, Elliott P, Elosua R, Eriksson JG, Freimer NB, Geus EJC, Glorioso N, Haiqing S, Hartikainen A-L, Havulinna AS, Hicks AA, Hui J, Igl W, Illig T, Jula A, Kajantie E, Kilpeläinen TO, Koiranen M, Kolcic I, Koskinen S, Kovacs P, Laitinen J, Liu J, Lokki M-L, Marusic A, Maschio A, Meitinger T, Mulas A, Paré G, Parker AN, Peden JF, Petersmann A, Pichler I, Pietiläinen KH, Pouta A, Ridderstråle M, Rotter JI, Sambrook JG, Sanders AR, Oliver Schmidt C, Sinisalo J, Smit JH, Stringham HM, Bragi Walters G, Widen E, Wild SH, Willemsen G, Zagato L, Zgaga L, Zitting P, Alavere H, Farrall M, McArdle WL, Nelis M, Peters MJ, Ripatti S, van Meurs JBJ, Aben KK, Ardlie KG, Beckmann JS, Beilby JP, Bergman RN, Bergmann S, Collins FS, Cusi D, den Heijer M, Eiriksdottir G, Gejman PV, Hall AS, Hamsten A, Huikuri HV, Iribarren C, Kähönen M, Kaprio J, Kathiresan S, Kiemeney L, Kocher T, Launer LJ, Lehtimäki T, Melander O, Mosley Jr TH, Musk AW, Nieminen MS, O'Donnell CJ, Ohlsson C, Oostra B, Palmer LJ, Raitakari O, Ridker PM, Rioux JD, Rissanen A, Rivolta C, Schunkert H, Shuldiner AR, Siscovick DS, Stumvoll M, Tönjes A, Tuomilehto J, van Ommen G-J, Viikari J, Heath AC, Martin NG,

Montgomery GW, Province MA, Kayser M, Arnold AM, Atwood LD, Boerwinkle E, Chanock SJ, Deloukas P, Gieger C, Grönberg H, Hall P, Hattersley AT, Hengstenberg C, Hoffman W, Mark Lathrop G, Salomaa V, Schreiber S, Uda M, Waterworth D, Wright AF, Assimes TL, Barroso I, Hofman A, Mohlke KL, Boomsma DI, Caulfield MJ, Adrienne Cupples L, Erdmann J, Fox CS, Gudnason V, Gyllensten U, Harris TB, Hayes RB, Jarvelin M-R, Mooser V, Munroe PB, Ouwehand WH, Penninx BW, Pramstaller PP, Quertermous T, Rudan I, Samani NJ, Spector TD, Völzke H, Watkins H, Wilson JF, Groop LC, Haritunians T, Hu FB, Kaplan RC, Metspalu A, North KE, Schlessinger D, Wareham NJ, Hunter DJ, O'Connell JR, Strachan DP, Wichmann HE, Borecki IB, van Duijn CM, Schadt EE, Thorsteinsdottir U, Peltonen L, Uitterlinden AG, Visscher PM, Chatterjee N, Loos RJF, Boehnke M, McCarthy MI, Ingelsson E, Lindgren CM, Abecasis GR, Stefansson K, Frayling TM, Hirschhorn JN. Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature. 2010;467:832.

15. Medland SE, Zayats T, Glaser B, Nyholt DR, Gordon SD, Wright MJ, Montgomery GW, Campbell MJ, Henders AK, Timpson NJ, Peltonen L, Wolke D, Ring SM, Deloukas P, Martin NG, Smith GD, Evans DM. A Variant in LIN28B Is Associated with 2D:4D Finger-Length Ratio, a Putative Retrospective Biomarker of Prenatal Testosterone Exposure. The American Journal of Human Genetics. 2010;86(4):519-25.

16. Cousminer DL, Berry DJ, Timpson NJ, Ang W, Thiering E, Byrne EM, Taal HR, Huikari V, Bradfield JP, Kerkhof M, Groen-Blokhuis MM, Kreiner-Møller E, Marinelli M, Holst C, Leinonen JT, Perry JRB, Surakka I, Pietiläinen O, Kettunen J, Anttila V, Kaakinen M, Sovio U, Pouta A, Das S, Lagou V, Power C, Prokopenko I, Evans DM, Kemp JP, St Pourcain B, Ring S, Palotie A, Kajantie E, Osmond C, Lehtimäki T, Viikari JS, Kähönen M, Warrington NM, Lye SJ, Palmer LJ, Tiesler CMT, Flexeder C, Montgomery GW, Medland SE, Hofman A, Hakonarson H, Guxens M, Bartels M, Salomaa V, Murabito JM, Kaprio J, Sørensen TIA, Ballester F, Bisgaard H, Boomsma DI, Koppelman GH, Grant SFA, Jaddoe VWV, Martin NG, Heinrich J, Pennell CE, Raitakari OT, Eriksson JG, Smith GD, Hyppönen E, Järvelin M-R, McCarthy MI, Ripatti S, Widén E. Genome-wide association and longitudinal analyses reveal genetic loci linking pubertal height growth, pubertal timing and childhood adiposity. Hum Mol Genet. 2013;22(13):2735-47.

17. Hyde CL, Nagle MW, Tian C, Chen X, Paciga SA, Wendland JR, Tung JY, Hinds DA, Perlis RH, Winslow AR. Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. Nat Genet. 2016;48(9):1031-6.

18. Leinonen JT, Surakka I, Havulinna AS, Kettunen J, Luoto R, Salomaa V, Widén E. Association of LIN28B with Adult Adiposity-Related Traits in Females. PLoS One. 2012;7(11):e48785.

19. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, Vukcevic D, Delaneau O, O'Connell J, Cortes A, Welsh S, Young A, Effingham M, McVean G, Leslie S, Allen N, Donnelly P, Marchini J. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018;562(7726):203-9.

20. Cannon ME, Mohlke KL. Deciphering the Emerging Complexities of Molecular Mechanisms at GWAS Loci. The American Journal of Human Genetics. 2018;103(5):637-53.

21. Stiles J, Jernigan TL. The basics of brain development. Neuropsychol Rev. 2010;20(4):327-48.

22. Swerdloff RS, Odell WD. Hormonal mechanisms in the onset of puberty. Postgrad Med J. 1975;51(594):200-8.

23. Reiter EO, Root AW. Hormonal Changes of Adolescence. Med Clin North Am. 1975;59(6):1289-304.

24. Herbison AE. Control of puberty onset and fertility by gonadotropin-releasing hormone neurons. Nature Reviews Endocrinology. 2016;12:452.

25. Plant TM. Neuroendocrine control of the onset of puberty. Front Neuroendocrinol. 2015;38:73-88.

26. Kuiri-Hänninen T, Sankilampi U, Dunkel L. Activation of the Hypothalamic-Pituitary-Gonadal Axis in Infancy: Minipuberty. Horm Res Paediatr. 2014;82(2):73-80.

27. Kaplan SL, Grumbach MM. THE ONTOGENESIS OF HUMAN FOETAL HORMONES. II. LUTEINIZING HORMONE (LH) AND FOLLICLE STIMULATING HORMONE (FSH). Acta Endocrinol. 1976;81(4):808.

28. Chevrier L, Delezoide AL, Guimiot F, de Roux N, Chevenne D, Dreux S, Caraty A. Negative Fetal FSH/LH Regulation in Late Pregnancy Is Associated with Declined Kisspeptin/KISS1R Expression in the Tuberal Hypothalamus. The Journal of Clinical Endocrinology & Metabolism. 2012;97(12):E2221-E9.

29. Machluf Y, Gutnick A, Levkowitz G. Development of the zebrafish hypothalamus. Ann N Y Acad Sci. 2011;1220(1):93-105.

30. Balasubramanian R, Dwyer A, Seminara SB, Pitteloud N, Kaiser UB, Crowley WF, Jr. Human GnRH deficiency: a unique disease model to unravel the ontogeny of GnRH neurons. Neuroendocrinology. 2010;92(2):81-99.

31. Dodé C, Hardelin J-PJJoMM. Kallmann syndrome: fibroblast growth factor signaling insufficiency? 2004;82(11):725-34.

32. Hardelin J-P. Kallmann syndrome: towards molecular pathogenesis. Mol Cell Endocrinol. 2001;179(1):75-81.

33. David K, Dingemanse E, Freud J, Laqueur E. Über krystallinisches männliches Hormon aus Hoden (Testosteron), wirksamer als aus Harn oder aus Cholesterin bereitetes Androsteron. Hoppe-Seyler´s Zeitschrift für physiologische Chemie1935. p. 281.

34. Tata JR. One hundred years of hormones. EMBO reports. 2005;6(6):490-6.
35. Matsuo H, Baba Y, Nair RMG, Arimura A, Schally AV. Structure of the porcine LH- and FSH-releasing hormone. I. The proposed amino acid sequence. Biochem Biophys Res Commun. 1971;43(6):1334-9.
36. Amoss M, Burgus R, Blackwell R, Vale W, Fellows R, Guillemin R. Purification, amino acid composition and N-terminus of the hypothalamic luteinizing hormone releasing factor (LRF) of ovine origin. Biochem Biophys Res Commun. 1971;44(1):205-10.
37. Ojeda SR, Dubay C, Lomniczi A, Kaidar G, Matagne V, Sandau US, Dissen GA. Gene networks and the neuroendocrine regulation of puberty. Mol Cell Endocrinol. 2010;324(1-2):3-11.
38. Sisk CL, Foster DL. The neural basis of puberty and adolescence. Nat Neurosci. 2004;7(10):1040-7.
39. de Roux N, Genin E, Carel J-C, Matsuda F, Chaussain J-L, Milgrom E. Hypogonadotropic hypogonadism due to loss of function of the KiSS1-derived peptide receptor GPR54. Proc Natl Acad Sci U S A. 2003;100(19):10972-6.
40. Seminara SB, Messager S, Chatzidaki EE, Thresher RR, Acierno JS, Shagoury JK, Bo-Abbas Y, Kuohung W, Schwinof KM, Hendrick AG, Zahn D, Dixon J, Kaiser UB, Slaugenhaupt SA, Gusella JF, O'Rahilly S, Carlton MBL, Crowley WF, Aparicio SAJR, Colledge WH. The GPR54 Gene as a Regulator of Puberty. N Engl J Med. 2003;349(17):1614-27.
41. Topaloglu AK, Reimann F, Guclu M, Yalin AS, Kotan LD, Porter KM, Serin A, Mungan NO, Cook JR, Ozbek MN, Imamoglu S, Akalin NS, Yuksel B, O'Rahilly S, Semple RK. TAC3 and TACR3 mutations in familial hypogonadotropic hypogonadism reveal a key role for Neurokinin B in the central control of reproduction. Nat Genet. 2008;41:354.
42. Ramaswamy S, Gibbs RB, Plant TM. Studies of the localisation of kisspeptin within the pituitary of the rhesus monkey (Macaca mulatta) and the effect of kisspeptin on the release of non-gonadotropic pituitary hormones. J Neuroendocrinol. 2009;21(10):795-804.
43. Ali B, Amin NA, Ramaswamy S, Plant TM, Seminara SB, Ciofi P. Neurokinin B Stimulates GnRH Release in the Male Monkey (Macaca mulatta) and Is Colocalized with Kisspeptin in the Arcuate Nucleus. Endocrinology. 2010;151(9):4494-503.
44. Terasawa E. Role of GABA in the mechanism of the onset of puberty in non-human primates. Int Rev Neurobiol. 2005;71:113-29.
45. Parent A-S, Matagne V, Bourguignon J-PJE. Control of puberty by excitatory amino acid neurotransmitters and its clinical implications. Endocrine. 2005;28(3):281-5.
46. Plant TM, Gay VL, Marshall GR, Arslan M. Puberty in monkeys is triggered by chemical stimulation of the hypothalamus. Proc Natl Acad Sci U S A 1989;86(7):2506-10.
47. Navarro VM, Gottsch ML, Chavkin C, Okamura H, Clifton DK, Steiner RA. Regulation of gonadotropin-releasing hormone secretion by kisspeptin/dynorphin/neurokinin B neurons in the arcuate nucleus of the mouse. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2009;29(38):11859-66.
48. Kaiser M, Lim S, Han CE, Uhlhaas PJ. Preferential Detachment During Human Brain Development: Age- and Sex-Specific Structural Connectivity in Diffusion Tensor Imaging (DTI) Data. Cereb Cortex. 2013;25(6):1477-89.
49. Onyango PO, Gesquiere LR, Altmann J, Alberts SC. Puberty and dispersal in a wild primate population. Horm Behav. 2013;64(2):240-9.
50. Barson NJ, Aykanat T, Hindar K, Baranski M, Bolstad GH, Fiske P, Jacq C, Jensen AJ, Johnston SE, Karlsson S, Kent M, Moen T, Niemelä E, Nome T, Næsje TF, Orell P, Romakkaniemi A, Sægrov H, Urdal K, Erkinaro J, Lien S, Primmer CR. Sex-dependent dominance at a single locus maintains variation in age at maturity in salmon. Nature. 2015;528:405.
51. Rosen DS. Physiologic Growth and Development During Adolescence. 2004;25(6):194-200.
52. Kaplowitz PB. Link Between Body Fat and the Timing of Puberty. Pediatrics. 2008;121(Supplement 3):S208-S17.
53. Marceau K, Ram N, Houts RM, Grimm KJ, Susman EJ. Individual differences in boys' and girls' timing and tempo of puberty: modeling development with nonlinear growth models. Dev Psychol. 2011;47(5):1389-409.
54. WARREN MP. The Effects of Exercise on Pubertal Progression and Reproductive Function in Girls*. The Journal of Clinical Endocrinology & Metabolism. 1980;51(5):1150-7.
55. Freedman DS, Khan LK, Serdula MK, Dietz WH, Srinivasan SR, Berenson GS, Bogalusa heart s. The relation of menarcheal age to obesity in childhood and adulthood: the Bogalusa heart study. BMC Pediatr. 2003;3:3-.
56. Lee JM, Kaciroti N, Appugliese D, Corwyn RF, Bradley RH, Lumeng JC. Body Mass Index and Timing of Pubertal Initiation in BoysBMI and Timing of Pubertal Initiation in Boys. Arch Pediatr Adolesc Med. 2010;164(2):139-44.
57. Yousefi M, Karmaus W, Zhang H, Roberts G, Matthews S, Clayton B, Arshad SHJWJoP. Relationships between age of puberty onset and height at age 18 years in girls and boys. World J Pediatr. 2013;9(3):230-8.

58. Biro FM, McMahon RP, Striegel-Moore R, Crawford PB, Obarzanek E, Morrison JA, Barton BA, Falkner F. Impact of timing of pubertal maturation on growth in black and white female adolescents: The National Heart, Lung, and Blood Institute Growth and Health Study. The Journal of Pediatrics. 2001;138(5):636-43.
59. van Dommelen P, de Kroon MLA, Cameron N, Schönbeck Y, van Buuren S. The impact of height during childhood on the national prevalence rates of overweight. PLoS One. 2014;9(1):e85769-e.
60. Stovitz SD, Demerath EW, Hannan PJ, Lytle LA, Himes JH. Growing into obesity: patterns of height growth in those who become normal weight, overweight, or obese as young adults. American journal of human biology : the official journal of the Human Biology Council. 2011;23(5):635-41.
61. Fretzayas A, Papadimitriou DT, Douros K, Nicolaidou P, Pantsiotou S, Papadimitriou A. Timing of Pubertal Onset in Girls: Evidence for Non-Gaussian Distribution. The Journal of Clinical Endocrinology & Metabolism. 2008;93(11):4422-5.
62. Marshall WA, Tanner JM. Variations in pattern of pubertal changes in girls. Arch Dis Child. 1969;44(235):291-303.
63. Marshall WA, Tanner JM. Variations in the pattern of pubertal changes in boys. Arch Dis Child. 1970;45(239):13-23.
64. Gajdos ZKZ, Henderson KD, Hirschhorn JN, Palmert MR. Genetic determinants of pubertal timing in the general population. Mol Cell Endocrinol. 2010;324(1-2):21-9.
65. Cousminer DL, Widén E, Palmert MR. The genetics of pubertal timing in the general population: recent advances and evidence for sex-specificity. Curr Opin Endocrinol Diabetes Obes. 2016;23(1):57-65.
66. Wehkalampi K, Silventoinen K, Kaprio J, Dick DM, Rose RJ, Pulkkinen L, Dunkel L. Genetic and environmental influences on pubertal timing assessed by height growth. American journal of human biology : the official journal of the Human Biology Council. 2008;20(4):417-23.
67. Wyshak G, Frisch RE. Evidence for a Secular Trend in Age of Menarche. N Engl J Med. 1982;306(17):1033-5.
68. KANTERO R-L, WIDHOLM O. THE AGE OF MENARCHE IN FINNISH GIRLS IN 1969. Obstet Gynecol Surv. 1972;27(8):632-3.
69. Herman-Giddens ME. The Decline in the Age of Menarche in the United States: Should We Be Concerned? J Adolesc Health. 2007;40(3):201-3.
70. Chen W, Ge W. Gonad differentiation and puberty onset in the zebrafish: Evidence for the dependence of puberty onset on body growth but not age in females. Mol Reprod Dev. 2013;80(5):384-92.
71. Jacobsen BK, Heuch I, Kvåle G. Association of Low Age at Menarche with Increased All-Cause Mortality: A 37-Year Follow-up of 61,319 Norwegian Women. Am J Epidemiol. 2007;166(12):1431-7.
72. Hayes FJ, Seminara SB, Crowley WF. HYPOGONADOTROPIC HYPOGONADISM. Endocrinol Metab Clin North Am. 1998;27(4):739-63.
73. Laitinen E-M, Vaaralahti K, Tommiska J, Eklund E, Tervaniemi M, Valanne L, Raivio TJOJoRD. Incidence, Phenotypic Features and Molecular Genetics of Kallmann Syndrome in Finland. Orphanet J Rare Dis. 2011;6(1):41.
74. Lomniczi A, Loche A, Castellano JM, Ronnekleiv OK, Bosch M, Kaidar G, Knoll JG, Wright H, Pfeifer GP, Ojeda SR. Epigenetic control of female puberty. Nat Neurosci. 2013;16(3):281-9.
75. Day FR, Perry JRB, Ong KK. Genetic Regulation of Puberty Timing in Humans. Neuroendocrinology. 2015;102(4):247-55.
76. Zhu J, Kusa TO, Chan Y-M. Genetics of pubertal timing. Curr Opin Pediatr. 2018;30(4):532-40.
77. Song JW, Hwang H-J, Lee CM, Park GH, Kim CS, Lee SJ, Ihm S-H. Hypogonadotrophic hypogonadism due to a mutation in the luteinizing hormone β-subunit gene. The Korean journal of internal medicine. 2018;33(3):638-41.
78. Weiss J, Axelrod L, Whitcomb RW, Harris PE, Crowley WF, Jameson JL. Hypogonadism Caused by a Single Amino Acid Substitution in the β Subunit of Luteinizing Hormone. N Engl J Med. 1992;326(3):179-83.
79. Desroches A, Beau I, Lahuna O, Misrahi M, Meduri G, Kuttenn F, Touraine P, Vacher-Lavenu MC. Delayed Puberty and Primary Amenorrhea Associated with a Novel Mutation of the Human Follicle-Stimulating Hormone Receptor: Clinical, Histological, and Molecular Studies. The Journal of Clinical Endocrinology & Metabolism. 2003;88(8):3491-8.
80. Aguirre RS, Eugster EA. Central precocious puberty: From genetics to treatment. Best Practice & Research Clinical Endocrinology & Metabolism. 2018;32(4):343-54.
81. Teles MG, Bianco SDC, Brito VN, Trarbach EB, Kuohung W, Xu S, Seminara SB, Mendonca BB, Kaiser UB, Latronico AC. A GPR54-Activating Mutation in a Patient with Central Precocious Puberty. N Engl J Med. 2008;358(7):709-15.
82. Latronico AC, Abreu AP, Silveira-Neto AP, Mendonca BB, Arnhold IJP, Santos MG, Brito VN, Silveira LG, Xu S, Noel SD, Bianco SDC, Kaiser UB, Kuohung W, Gryngarten M, Escobar ME. Mutations of the KISS1 Gene in Disorders of Puberty. The Journal of Clinical Endocrinology & Metabolism. 2010;95(5):2276-80.

83. Abreu AP, Dauber A, Macedo DB, Noel SD, Brito VN, Gill JC, Cukier P, Thompson IR, Navarro VM, Gagliardi PC, Rodrigues T, Kochi C, Longui CA, Beckers D, de Zegher F, Montenegro LR, Mendonca BB, Carroll RS, Hirschhorn JN, Latronico AC, Kaiser UB. Central Precocious Puberty Caused by Mutations in the Imprinted Gene MKRN3. N Engl J Med. 2013;368(26):2467-75.
84. Andrew M, Dauber A, Weirauch MT, Labilloy G, Latronico AC, Mendonca BB, Bessa DS, Macedo DB, Montenegro LR, Cunha-Silva M, Brito VN, Abreu AP, Carroll RS, Kaiser UB, Roberts SA, Kirby A, Jacobs DC, Chappell PE, Haig D. Paternally Inherited DLK1 Deletion Associated With Familial Central Precocious Puberty. The Journal of Clinical Endocrinology & Metabolism. 2017;102(5):1557-67.
85. McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JPA, Hirschhorn JN. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nature Reviews Genetics. 2008;9:356.
86. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, Yang J. 10 Years of GWAS Discovery: Biology, Function, and Translation. The American Journal of Human Genetics. 2017;101(1):5-22.
87. Storey JD, Tibshirani R. Statistical significance for genomewide studies. Proc Natl Acad Sci U S A. 2003;100(16):9440-5.
88. Klein RJ, Zeiss C, Chew EY, Tsai J-Y, Sackler RS, Haynes C, Henning AK, SanGiovanni JP, Mane SM, Mayne ST, Bracken MB, Ferris FL, Ott J, Barnstable C, Hoh J. Complement factor H polymorphism in age-related macular degeneration. Science (New York, NY). 2005;308(5720):385-9.
89. Ong KK, Elks CE, Li S, Zhao JH, Luan Ja, Andersen LB, Bingham SA, Brage S, Smith GD, Ekelund U, Gillson CJ, Glaser B, Golding J, Hardy R, Khaw K-T, Kuh D, Luben R, Marcus M, McGeehin MA, Ness AR, Northstone K, Ring SM, Rubin C, Sims MA, Song K, Strachan DP, Vollenweider P, Waeber G, Waterworth DM, Wong A, Deloukas P, Barroso I, Mooser V, Loos RJ, Wareham NJ. Genetic variation in LIN28B is associated with the timing of puberty. Nat Genet. 2009;41(6):729-33.
90. Sulem P, Gudbjartsson DF, Rafnar T, Holm H, Olafsdottir EJ, Olafsdottir GH, Jonsson T, Alexandersen P, Feenstra B, Boyd HA, Aben KK, Verbeek ALM, Roeleveld N, Jonasdottir A, Styrkarsdottir U, Steinthorsdottir V, Karason A, Stacey SN, Gudmundsson J, Jakobsdottir M, Thorleifsson G, Hardarson G, Gulcher J, Kong A, Kiemeney LA, Melbye M, Christiansen C, Tryggvadottir L, Thorsteinsdottir U, Stefansson K. Genome-wide association study identifies sequence variants on 6q21 associated with age at menarche. Nat Genet. 2009;41:734.
91. Perry JRB, Stolk L, Franceschini N, Lunetta KL, Zhai G, McArdle PF, Smith AV, Aspelund T, Bandinelli S, Boerwinkle E, Cherkas L, Eiriksdottir G, Estrada K, Ferrucci L, Folsom AR, Garcia M, Gudnason V, Hofman A, Karasik D, Kiel DP, Launer LJ, van Meurs J, Nalls MA, Rivadeneira F, Shuldiner AR, Singleton A, Soranzo N, Tanaka T, Visser JA, Weedon MN, Wilson SG, Zhuang V, Streeten EA, Harris TB, Murray A, Spector TD, Demerath EW, Uitterlinden AG, Murabito JM. Meta-analysis of genome-wide association data identifies two loci influencing age at menarche. Nat Genet. 2009;41(6):648-50.
92. He C, Kraft P, Chen C, Buring JE, Paré G, Hankinson SE, Chanock SJ, Ridker PM, Hunter DJ, Chasman DI. Genome-wide association studies identify loci associated with age at menarche and age at natural menopause. Nat Genet. 2009;41(6):724-8.
93. Consortium ftEGG, Consortium tR, Cousminer DL, Widén E, Ripatti S, Surakka I, Kettunen J, Körner A, Neef M, Weise S, Kiess W, Siitonen N, Pahkala K, Niinikoski H, Raitakari OT, Dedoussis GV, Ntalla I, Marinelli M, Bustamante M, Torrent M, Perry JRB, Kähönen M, Viikari J, Lehtimäki T, Hulshoff Pol HE, Zeggini E, Panoutsopoulou K, Murabito J, Stergiakouli E, Davey Smith G, Timpson NJ, McCarthy MI, Mcmahon G, Ring S, Power C, Berry DJ, Hyppönen E, Wang C, Pennell CE, Ang W, Groen-Blokhuis MM, Boomsma DI, Middeldorp CM, Penninx BWJH, Jansen R. Genome-wide association study of sexual maturation in males and females highlights a role for body mass and menarche loci in male puberty. Hum Mol Genet. 2014;23(16):4452-64.
94. Day FR, Bulik-Sullivan B, Hinds DA, Finucane HK, Murabito JM, Tung JY, Ong KK, Perry JRB. Shared genetic aetiology of puberty timing between sexes and with health-related outcomes. Nature communications. 2015;6:8842-.
95. Cousminer DL, Stergiakouli E, Berry DJ, Ang W, Groen-Blokhuis MM, Körner A, Siitonen N, Ntalla I, Marinelli M, Perry JRB, Kettunen J, Jansen R, Surakka I, Timpson NJ, Ring S, McMahon G, Power C, Wang C, Kähönen M, Viikari J, Lehtimäki T, Middeldorp CM, Hulshoff Pol HE, Neef M, Weise S, Pahkala K, Niinikoski H, Zeggini E, Panoutsopoulou K, Bustamante M, Penninx BWJH, ReproGen C, Murabito J, Torrent M, Dedoussis GV, Kiess W, Boomsma DI, Pennell CE, Raitakari OT, Hyppönen E, Davey Smith G, Ripatti S, McCarthy MI, Widén E, Early Growth Genetics C. Genome-wide association study of sexual maturation in males and females highlights a role for body mass and menarche loci in male puberty. Hum Mol Genet. 2014;23(16):4452-64.
96. Chuang JC, Jones PA. Epigenetics and MicroRNAs. Pediatr Res. 2007;61:24R.
97. Moutinho C, Esteller M. Chapter Seven - MicroRNAs and Epigenetics. In: Croce CM, Fisher PB, editors. Adv Cancer Res. 135: Academic Press; 2017. p. 189-220.
98. Curley JP, Mashoodh R, Champagne FA. Epigenetics and the origins of paternal effects. Horm Behav. 2011;59(3):306-14.

99. Barbaux S, Gascoin-Lachambre G, Buffat C, Monnier P, Mondon F, Tonanny M-B, Pinard A, Auer J, Bessières B, Barlier A, Jacques S, Simeoni U, Dandolo L, Letourneur F, Jammes H, Vaiman D. A genome-wide approach reveals novel imprinted genes expressed in the human placenta. Epigenetics. 2012;7(9):1079-90.
100. Gibney ER, Nolan CM. Epigenetics and gene expression. Heredity (Edinb). 2010;105:4.
101. Toro CA, Wright H, Aylwin CF, Ojeda SR, Lomniczi A. Trithorax dependent changes in chromatin landscape at enhancer and promoter regions drive female puberty. Nature Communications. 2018;9(1):57.
102. Morrison KE, Rodgers AB, Morgan CP, Bale TL. Epigenetic mechanisms in pubertal brain maturation. Neuroscience. 2014;264:17-24.
103. Guo Y, Chen Y, Ito H, Watanabe A, Ge X, Kodama T, Aburatani H. Identification and characterization of lin-28 homolog B (LIN28B) in human hepatocellular carcinoma. Gene. 2006;384:51-61.
104. Moss EG, Lee RC, Ambros V. The Cold Shock Domain Protein LIN-28 Controls Developmental Timing in C. elegans and Is Regulated by the lin-4 RNA. Cell. 1997;88(5):637-46.
105. Yokoyama S, Hashimoto M, Shimizu H, Ueno-Kudoh H, Uchibe K, Kimura I, Asahara H. Dynamic gene expression of Lin-28 during embryonic development in mouse and chicken. Gene Expression Patterns. 2008;8(3):155-60.
106. Moss EG, Tang L. Conservation of the heterochronic regulator Lin-28, its developmental expression and microRNA complementary sites. Dev Biol. 2003;258(2):432-42.
107. Viswanathan SR, Daley GQ. Lin28: A MicroRNA Regulator with a Macro Role. Cell. 2010;140(4):445-9.
108. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler, David. The Human Genome Browser at UCSC. Genome Res. 2002;12(6):996-1006.
109. Chalfie M, Horvitz HR, Sulston JE. Mutations that lead to reiterations in the cell lineages of C. elegans. Cell. 1981;24(1):59-69.
110. Lee RC, Feinbaum RL, Ambros V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell. 1993;75(5):843-54.
111. Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, Rougvie AE, Horvitz HR, Ruvkun G. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature. 2000;403(6772):901-6.
112. Morita K, Han M. Multiple mechanisms are involved in regulating the expression of the developmental timing regulator lin-28 in Caenorhabditis elegans. EMBO J. 2006;25(24):5794-804.
113. Shyh-Chang N, Daley George Q. Lin28: Primal Regulator of Growth and Metabolism in Stem Cells. Cell Stem Cell. 2013;12(4):395-406.
114. Pasquinelli AE, Reinhart BJ, Slack F, Martindale MQ, Kuroda MI, Maller B, Hayward DC, Ball EE, Degnan B, Müller P, Spring J, Srinivasan A, Fishman M, Finnerty J, Corbo J, Levine M, Leahy P, Davidson E, Ruvkun G. Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. Nature. 2000;408(6808):86-9.
115. Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, Frane JL, Tian S, Nie J, Jonsdottir GA, Ruotti V, Stewart R, Slukvin II, Thomson JA. Induced Pluripotent Stem Cell Lines Derived from Human Somatic Cells. Science. 2007;318(5858):1917-20.
116. Zhang J, Ratanasirintrawoot S, Chandrasekaran S, Wu Z, Ficarro Scott B, Yu C, Ross Christian A, Cacchiarelli D, Xia Q, Seligson M, Shinoda G, Xie W, Cahan P, Wang L, Ng S-C, Tintara S, Trapnell C, Onder T, Loh Y-H, Mikkelsen T, Sliz P, Teitell Michael A, Asara John M, Marto Jarrod A, Li H, Collins James J, Daley George Q. LIN28 Regulates Stem Cell Metabolism and Conversion to Primed Pluripotency. Cell Stem Cell. 2016;19(1):66-80.
117. Heo I, Joo C, Cho J, Ha M, Han J, Kim VN. Lin28 Mediates the Terminal Uridylation of let-7 Precursor MicroRNA. Mol Cell. 2008;32(2):276-84.
118. Piskounova E, Polytarchou C, Thornton James E, LaPierre Robert J, Pothoulakis C, Hagan John P, Iliopoulos D, Gregory Richard I. Lin28A and Lin28B Inhibit let-7 MicroRNA Biogenesis by Distinct Mechanisms. Cell. 2011;147(5):1066-79.
119. Hagan JP, Piskounova E, Gregory RI. Lin28 recruits the TUTase Zcchc11 to inhibit let-7 maturation in mouse embryonic stem cells. Nature Structural &Amp; Molecular Biology. 2009;16:1021.
120. Rybak A, Fuchs H, Smirnova L, Brandt C, Pohl EE, Nitsch R, Wulczyn FG. A feedback loop comprising lin-28 and let-7 controls pre-let-7 maturation during neural stem-cell commitment. Nat Cell Biol. 2008;10:987.
121. Viswanathan SR, Daley GQ, Gregory RI. Selective Blockade of MicroRNA Processing by Lin28. Science. 2008;320(5872):97-100.
122. Rahkonen N, Stubb A, Malonzo M, Edelman S, Emani MR, Närvä E, Lähdesmäki H, Ruohola-Baker H, Lahesmaa R, Lund R. Mature Let-7 miRNAs fine tune expression of LIN28B in pluripotent human embryonic stem cells. Stem Cell Research. 2016;17(3):498-503.
123. Balzeau J, Menezes MR, Cao S, Hagan JP. The LIN28/let-7 Pathway in Cancer. Frontiers in genetics. 2017;8:31-.
124. Thornton JE, Gregory RI. How does Lin28 let-7 control development and disease? Trends Cell Biol. 2012;22(9):474-82.

125. Molenaar JJ, Domingo-Fernandez R, Ebus ME, Lindner S, Koster J, Drabek K, Mestdagh P, van Sluis P, Valentijn LJ, van Nes J, Broekmans M, Haneveld F, Volckmann R, Bray I, Heukamp L, Sprussel A, Thor T, Kieckbusch K, Klein-Hitpass L, Fischer M, Vandesompele J, Schramm A, van Noesel MM, Varesio L, Speleman F, Eggert A, Stallings RL, Caron HN, Versteeg R, Schulte JH. LIN28B induces neuroblastoma and enhances MYCN levels via let-7 suppression. Nat Genet. 2012;44(11):1199-206.
126. Yang J, Bennett BD, Luo S, Inoue K, Grimm SA, Schroth GP, Bushel PR, Kinyamu HK, Archer TK. LIN28A Modulates Splicing and Gene Expression Programs in Breast Cancer Cells. Mol Cell Biol. 2015;35(18):3225-43.
127. Tummala R, Nadiminty N, Lou W, Zhu Y, Gandour-Edwards R, Chen H-W, Evans CP, Gao AC. Lin28 Promotes Growth of Prostate Cancer Cells and Activates the Androgen Receptor. The American Journal of Pathology. 2013;183(1):288-95.
128. Wang T, Wang G, Hao D, Liu X, Wang D, Ning N, Li XJMC. Aberrant regulation of the LIN28A/LIN28B and let-7 loop in human malignant tumors and its effects on the hallmarks of cancer. Mol Cancer. 2015;14(1):125.
129. Zhu H, Shah S, Shyh-Chang N, Shinoda G, Einhorn WS, Viswanathan SR, Takeuchi A, Grasemann C, Rinn JL, Lopez MF, Hirschhorn JN, Palmert MR, Daley GQ. Lin28a transgenic mice manifest size and puberty phenotypes identified in human genetic association studies. Nat Genet. 2010;42(7):626-30.
130. Zhu H, Shyh-Chang N, Segrè Ayellet V, Shinoda G, Shah Samar P, Einhorn William S, Takeuchi A, Engreitz Jesse M, Hagan John P, Kharas Michael G, Urbach A, Thornton James E, Triboulet R, Gregory Richard I, Altshuler D, Daley George Q. The Lin28/let-7 Axis Regulates Glucose Metabolism. Cell. 2011;147(1):81-94.
131. Shinoda G, Shyh-Chang N, Soysa TYd, Zhu H, Seligson MT, Shah SP, Abo-Sido N, Yabuuchi A, Hagan JP, Gregory RI, Asara JM, Cantley LC, Moss EG, Daley GQ. Fetal Deficiency of Lin28 Programs Life-Long Aberrations in Growth and Glucose Metabolism. Stem Cells. 2013;31(8):1563-73.
132. Shyh-Chang N, Zhu H, Yvanka de Soysa T, Shinoda G, Seligson MT, Tsanov KM, Nguyen L, Asara JM, Cantley LC, Daley GQ. Lin28 enhances tissue repair by reprogramming cellular metabolism. Cell. 2013;155(4):778-92.
133. Corre C, Shinoda G, Zhu H, Cousminer DL, Crossman C, Bellissimo C, Goldenberg A, Daley GQ, Palmert MR. Sex-specific regulation of weight and puberty by the Lin28/let-7 axis. The Journal of endocrinology. 2016;228(3):179-91.
134. Sangiao-Alvarellos S, Manfredi-Lozano M, Ruiz-Pino F, Navarro VM, Sánchez-Garrido MA, Leon S, Dieguez C, Cordido F, Matagne V, Dissen GA, Ojeda SR, Pinilla L, Tena-Sempere M. Changes in Hypothalamic Expression of the Lin28/let-7 System and Related MicroRNAs During Postnatal Maturation and After Experimental Manipulations of Puberty. Endocrinology. 2013;154(2):942-55.
135. Faas L, Warrander FC, Maguire R, Ramsbottom SA, Quinn D, Genever P, Isaacs HV. Lin28 proteins are required for germ layer specification in Xenopus. Development. 2013;140(5):976-86.
136. Ouchi Y, Yamamoto J, Iwamoto T. The Heterochronic Genes lin-28a and lin-28b Play an Essential and Evolutionarily Conserved Role in Early Zebrafish Development. PLoS One. 2014;9(2):e88086.
137. Stratoulias V, Heino TI, Michon F. Lin-28 Regulates Oogenesis and Muscle Formation in Drosophila melanogaster. PLoS One. 2014;9(6):e101141.
138. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP, Gauthier LD, Brand H, Solomonson M, Watts NA, Rhodes D, Singer-Berk M, Seaby EG, Kosmicki JA, Walters RK, Tashman K, Farjoun Y, Banks E, Poterba T, Wang A, Seed C, Whiffin N, Chong JX, Samocha KE, Pierce-Hoffman E, Zappala Z, O'Donnell-Luria AH, Vallabh Minikel E, Weisburd B, Lek M, Ware JS, Vittal C, Armean IM, Bergelson L, Cibulskis K, Connolly KM, Covarrubias M, Donnelly S, Ferriera S, Gabriel S, Gentry J, Gupta N, Jeandet T, Kaplan D, Llanwarne C, Munshi R, Novod S, Petrillo N, Roazen D, Ruano-Rubio V, Saltzman A, Schleicher M, Soto J, Tibbetts K, Tolonen C, Wade G, Talkowski ME, Neale BM, Daly MJ, MacArthur DG. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. bioRxiv. 2019:531210.
139. Lettre G, Jackson AU, Gieger C, Schumacher FR, Berndt SI, Sanna S, Eyheramendy S, Voight BF, Butler JL, Guiducci C, Illig T, Hackett R, Heid IM, Jacobs KB, Lyssenko V, Uda M, Boehnke M, Chanock SJ, Groop LC, Hu FB, Isomaa B, Kraft P, Peltonen L, Salomaa V, Schlessinger D, Hunter DJ, Hayes RB, Abecasis GR, Wichmann HE, Mohlke KL, Hirschhorn JN. Identification of ten loci associated with height highlights new biological pathways in human growth. Nat Genet. 2008;40(5):584-91.
140. Morris JA, Kemp JP, Youlten SE, Laurent L, Logan JG, Chai RC, Vulpescu NA, Forgetta V, Kleinman A, Mohanty ST, Sergio CM, Quinn J, Nguyen-Yamamoto L, Luco A-L, Vijay J, Simon M-M, Pramatarova A, Medina-Gomez C, Trajanoska K, Ghirardello EJ, Butterfield NC, Curry KF, Leitch VD, Sparkes PC, Adoum A-T, Mannan NS, Komla-Ebri DSK, Pollard AS, Dewhurst HF, Hassall TAD, Beltejar M-JG, Agee M, Alipanahi B, Auton A, Bell RK, Bryc K, Elson SL, Fontanillas P, Furlotte NA, McCreight JC, Huber KE, Litterman NK, McIntyre MH, Mountain JL, Noblin ES, Northover CAM, Pitts SJ, Sathirapongsasuti JF, Sazonova OV, Shelton JF, Shringarpure S, Tian C, Tung JY, Vacic V, Wilson CH, Adams DJ, Vaillancourt SM, Kaptoge S, Baldock P, Cooper C, Reeve J, Ntzani EE, Evangelou E, Ohlsson C, Karasik D, Rivadeneira F, Kiel DP, Tobias JH, Gregson CL, Harvey NC, Grundberg E, Goltzman D, Adams DJ, Lelliott CJ,

Hinds DA, Ackert-Bicknell CL, Hsu Y-H, Maurano MT, Croucher PI, Williams GR, Bassett JHD, Evans DM, Richards JB, andMe Research T. An atlas of genetic influences on osteoporosis in humans and mice. Nat Genet. 2019;51(2):258-66.
141. Jansen PR, Watanabe K, Stringer S, Skene N, Bryois J, Hammerschlag AR, de Leeuw CA, Benjamins JS, Muñoz-Manchado AB, Nagel M, Savage JE, Tiemeier H, White T, Agee M, Alipanahi B, Auton A, Bell RK, Bryc K, Elson SL, Fontanillas P, Furlotte NA, Hinds DA, Huber KE, Kleinman A, Litterman NK, McCreight JC, McIntyre MH, Mountain JL, Noblin ES, Northover CAM, Pitts SJ, Sathirapongsasuti JF, Sazonova OV, Shelton JF, Shringarpure S, Tian C, Wilson CH, Tung JY, Hinds DA, Vacic V, Wang X, Sullivan PF, van der Sluis S, Polderman TJC, Smit AB, Hjerling-Leffler J, Van Someren EJW, Posthuma D, The 23andMe Research T. Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. Nat Genet. 2019;51(3):394-403.
142. Kim SK. Identification of 613 new loci associated with heel bone mineral density and a polygenic risk score for bone mineral density, osteoporosis and fracture. PLoS One. 2018;13(7):e0200785.
143. Nannini DR, Fan F, Kim H, Huang H, Gao XR. Genome-wide association analyses identify new loci influencing intraocular pressure. Hum Mol Genet. 2018;27(12):2205-13.
144. Shungin D, Winkler TW, Croteau-Chonka DC, Ferreira T, Locke AE, Mägi R, Strawbridge RJ, Pers TH, Fischer K, Justice AE, Workalemahu T, Wu JMW, Buchkovich ML, Heard-Costa NL, Roman TS, Drong AW, Song C, Gustafsson S, Day FR, Esko T, Fall T, Kutalik Z, Luan Ja, Randall JC, Scherag A, Vedantam S, Wood AR, Chen J, Fehrmann R, Karjalainen J, Kahali B, Liu C-T, Schmidt EM, Absher D, Amin N, Anderson D, Beekman M, Bragg-Gresham JL, Buyske S, Demirkan A, Ehret GB, Feitosa MF, Goel A, Jackson AU, Johnson T, Kleber ME, Kristiansson K, Mangino M, Mateo Leach I, Medina-Gomez C, Palmer CD, Pasko D, Pechlivanis S, Peters MJ, Prokopenko I, Stančáková A, Ju Sung Y, Tanaka T, Teumer A, Van Vliet-Ostaptchouk JV, Yengo L, Zhang W, Albrecht E, Ärnlöv J, Arscott GM, Bandinelli S, Barrett A, Bellis C, Bennett AJ, Berne C, Blüher M, Böhringer S, Bonnet F, Böttcher Y, Bruinenberg M, Carba DB, Caspersen IH, Clarke R, Warwick Daw E, Deelen J, Deelman E, Delgado G, Doney ASF, Eklund N, Erdos MR, Estrada K, Eury E, Friedrich N, Garcia ME, Giedraitis V, Gigante B, Go AS, Golay A, Grallert H, Grammer TB, Gräßler J, Grewal J, Groves CJ, Haller T, Hallmans G, Hartman CA, Hassinen M, Hayward C, Heikkilä K, Herzig K-H, Helmer Q, Hillege HL, Holmen O, Hunt SC, Isaacs A, Ittermann T, James AL, Johansson I, Juliusdottir T, Kalafati I-P, Kinnunen L, Koenig W, Kooner IK, Kratzer W, Lamina C, Leander K, Lee NR, Lichtner P, Lind L, Lindström J, Lobbens S, Lorentzon M, Mach F, Magnusson PKE, Mahajan A, McArdle WL, Menni C, Merger S, Mihailov E, Milani L, Mills R, Moayyeri A, Monda KL, Mooijaart SP, Mühleisen TW, Mulas A, Müller G, Müller-Nurasyid M, Nagaraja R, Nalls MA, Narisu N, Glorioso N, Nolte IM, Olden M, Rayner NW, Renstrom F, Ried JS, Robertson NR, Rose LM, Sanna S, Scharnagl H, Scholtens S, Sennblad B, Seufferlein T, Sitlani CM, Vernon Smith A, Stirrups K, Stringham HM, Sundström J, Swertz MA, Swift AJ, Syvänen A-C, Tayo BO, Thorand B, Thorleifsson G, Tomaschitz A, Troffa C, van Oort FVA, Verweij N, Vonk JM, Waite LL, Wennauer R, Wilsgaard T, Wojczynski MK, Wong A, Zhang Q, Hua Zhao J, Brennan EP, Choi M, Eriksson P, Folkersen L, Franco-Cereceda A, Gharavi AG, Hedman ÅK, Hivert M-F, Huang J, Kanoni S, Karpe F, Keildson S, Kiryluk K, Liang L, Lifton RP, Ma B, McKnight AJ, McPherson R, Metspalu A, Min JL, Moffatt MF, Montgomery GW, Murabito JM, Nicholson G, Nyholt DR, Olsson C, Perry JRB, Reinmaa E, Salem RM, Sandholm N, Schadt EE, Scott RA, Stolk L, Vallejo EE, Westra H-J, Zondervan KT, The AC, The CDC, The CC, The GC, The GC, The G, The I, The International Endogene C, The LifeLines Cohort S, The MI, The Mu TC, The PC, The ReproGen C, Amouyel P, Arveiler D, Bakker SJL, Beilby J, Bergman RN, Blangero J, Brown MJ, Burnier M, Campbell H, Chakravarti A, Chines PS, Claudi-Boehm S, Collins FS, Crawford DC, Danesh J, de Faire U, de Geus EJC, Dörr M, Erbel R, Eriksson JG, Farrall M, Ferrannini E, Ferrières J, Forouhi NG, Forrester T, Franco OH, Gansevoort RT, Gieger C, Gudnason V, Haiman CA, Harris TB, Hattersley AT, Heliövaara M, Hicks AA, Hingorani AD, Hoffmann W, Hofman A, Homuth G, Humphries SE, Hyppönen E, Illig T, Jarvelin M-R, Johansen B, Jousilahti P, Jula AM, Kaprio J, Kee F, Keinanen-Kiukaanniemi SM, Kooner JS, Kooperberg C, Kovacs P, Kraja AT, Kumari M, Kuulasmaa K, Kuusisto J, Lakka TA, Langenberg C, Le Marchand L, Lehtimäki T, Lyssenko V, Männistö S, Marette A, Matise TC, McKenzie CA, McKnight B, Musk AW, Möhlenkamp S, Morris AD, Nelis M, Ohlsson C, Oldehinkel AJ, Ong KK, Palmer LJ, Penninx BW, Peters A, Pramstaller PP, Raitakari OT, Rankinen T, Rao DC, Rice TK, Ridker PM, Ritchie MD, Rudan I, Salomaa V, Samani NJ, Saramies J, Sarzynski MA, Schwarz PEH, Shuldiner AR, Staessen JA, Steinthorsdottir V, Stolk RP, Strauch K, Tönjes A, Tremblay A, Tremoli E, Vohl M-C, Völker U, Vollenweider P, Wilson JF, Witteman JC, Adair LS, Bochud M, Boehm BO, Bornstein SR, Bouchard C, Cauchi S, Caulfield MJ, Chambers JC, Chasman DI, Cooper RS, Dedoussis G, Ferrucci L, Froguel P, Grabe H-J, Hamsten A, Hui J, Hveem K, Jöckel K-H, Kivimaki M, Kuh D, Laakso M, Liu Y, März W, Munroe PB, Njølstad I, Oostra BA, Palmer CNA, Pedersen NL, Perola M, Pérusse L, Peters U, Power C, Quertermous T, Rauramaa R, Rivadeneira F, Saaristo TE, Saleheen D, Sinisalo J, Eline Slagboom P, Snieder H, Spector TD, Thorsteinsdottir U, Stumvoll M, Tuomilehto J, Uitterlinden AG, Uusitupa M, van der Harst P, Veronesi G, Walker M, Wareham NJ, Watkins H, Wichmann HE, Abecasis GR, Assimes TL, Berndt SI, Boehnke M, Borecki IB, Deloukas P, Franke L, Frayling TM, Groop LC, Hunter DJ, Kaplan RC, O'Connell JR, Qi L, Schlessinger D, Strachan DP, Stefansson K, van Duijn CM, Willer CJ, Visscher PM, Yang J, Hirschhorn JN, Carola Zillikens M, McCarthy MI, Speliotes EK, North KE, Fox CS, Barroso I, Franks PW, Ingelsson E, Heid IM, Loos RJF, Cupples LA, Morris AP, Lindgren CM, Mohlke KL. New genetic loci link adipose and insulin biology to body fat distribution. Nature. 2015;518:187.

145. Graff M, Scott RA, Justice AE, Young KL, Feitosa MF, Barata L, Winkler TW, Chu AY, Mahajan A, Hadley D, Xue L, Workalemahu T, Heard-Costa NL, den Hoed M, Ahluwalia TS, Qi Q, Ngwa JS, Renström F, Quaye L, Eicher JD, Hayes JE, Cornelis M, Kutalik Z, Lim E, Luan Ja, Huffman JE, Zhang W, Zhao W, Griffin PJ, Haller T, Ahmad S, Marques-Vidal PM, Bien S, Yengo L, Teumer A, Smith AV, Kumari M, Harder MN, Justesen JM, Kleber ME, Hollensted M, Lohman K, Rivera NV, Whitfield JB, Zhao JH, Stringham HM, Lyytikäinen L-P, Huppertz C, Willemsen G, Peyrot WJ, Wu Y, Kristiansson K, Demirkan A, Fornage M, Hassinen M, Bielak LF, Cadby G, Tanaka T, Mägi R, van der Most PJ, Jackson AU, Bragg-Gresham JL, Vitart V, Marten J, Navarro P, Bellis C, Pasko D, Johansson Å, Snitker S, Cheng Y-C, Eriksson J, Lim U, Aadahl M, Adair LS, Amin N, Balkau B, Auvinen J, Beilby J, Bergman RN, Bergmann S, Bertoni AG, Blangero J, Bonnefond A, Bonnycastle LL, Borja JB, Brage S, Busonero F, Buyske S, Campbell H, Chines PS, Collins FS, Corre T, Smith GD, Delgado GE, Dueker N, Dörr M, Ebeling T, Eiriksdottir G, Esko T, Faul JD, Fu M, Færch K, Gieger C, Gläser S, Gong J, Gordon-Larsen P, Grallert H, Grammer TB, Grarup N, van Grootheest G, Harald K, Hastie ND, Havulinna AS, Hernandez D, Hindorff L, Hocking LJ, Holmens OL, Holzapfel C, Hottenga JJ, Huang J, Huang T, Hui J, Huth C, Hutri-Kähönen N, James AL, Jansson J-O, Jhun MA, Juonala M, Kinnunen L, Koistinen HA, Kolcic I, Komulainen P, Kuusisto J, Kvaløy K, Kähönen M, Lakka TA, Launer LJ, Lehne B, Lindgren CM, Lorentzon M, Luben R, Marre M, Milaneschi Y, Monda KL, Montgomery GW, De Moor MHM, Mulas A, Müller-Nurasyid M, Musk AW, Männikkö R, Männistö S, Narisu N, Nauck M, Nettleton JA, Nolte IM, Oldehinkel AJ, Olden M, Ong KK, Padmanabhan S, Paternoster L, Perez J, Perola M, Peters A, Peters U, Peyser PA, Prokopenko I, Puolijoki H, Raitakari OT, Rankinen T, Rasmussen-Torvik LJ, Rawal R, Ridker PM, Rose LM, Rudan I, Sarti C, Sarzynski MA, Savonen K, Scott WR, Sanna S, Shuldiner AR, Sidney S, Silbernagel G, Smith BH, Smith JA, Snieder H, Stančáková A, Sternfeld B, Swift AJ, Tammelin T, Tan S-T, Thorand B, Thuillier D, Vandenput L, Vestergaard H, van Vliet-Ostaptchouk JV, Vohl M-C, Völker U, Waeber G, Walker M, Wild S, Wong A, Wright AF, Zillikens MC, Zubair N, Haiman CA, Lemarchand L, Gyllensten U, Ohlsson C, Hofman A, Rivadeneira F, Uitterlinden AG, Pérusse L, Wilson JF, Hayward C, Polasek O, Cucca F, Hveem K, Hartman CA, Tönjes A, Bandinelli S, Palmer LJ, Kardia SLR, Rauramaa R, Sørensen TIA, Tuomilehto J, Salomaa V, Penninx BWJH, de Geus EJC, Boomsma DI, Lehtimäki T, Mangino M, Laakso M, Bouchard C, Martin NG, Kuh D, Liu Y, Linneberg A, März W, Strauch K, Kivimäki M, Harris TB, Gudnason V, Völzke H, Qi L, Järvelin M-R, Chambers JC, Kooner JS, Froguel P, Kooperberg C, Vollenweider P, Hallmans G, Hansen T, Pedersen O, Metspalu A, Wareham NJ, Langenberg C, Weir DR, Porteous DJ, Boerwinkle E, Chasman DI, Consortium C, Consortium EP-I, Consortium P, Abecasis GR, Barroso I, McCarthy MI, Frayling TM, O'Connell JR, van Duijn CM, Boehnke M, Heid IM, Mohlke KL, Strachan DP, Fox CS, Liu C-T, Hirschhorn JN, Klein RJ, Johnson AD, Borecki IB, Franks PW, North KE, Cupples LA, Loos RJF, Kilpeläinen TO. Genome-wide physical activity interactions in adiposity — A meta-analysis of 200,452 adults. PLOS Genetics. 2017;13(4):e1006528.

146. Diskin SJ, Capasso M, Schnepp RW, Cole KA, Attiyeh EF, Hou C, Diamond M, Carpenter EL, Winter C, Lee H, Jagannathan J, Latorre V, Iolascon A, Hakonarson H, Devoto M, Maris JM. Common variation at 6q16 within HACE1 and LIN28B influences susceptibility to neuroblastoma. Nat Genet. 2012;44:1126.

147. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, Tukiainen T, Birnbaum DP, Kosmicki JA, Duncan LE, Estrada K, Zhao F, Zou J, Pierce-Hoffman E, Berghout J, Cooper DN, Deflaux N, DePristo M, Do R, Flannick J, Fromer M, Gauthier L, Goldstein J, Gupta N, Howrigan D, Kiezun A, Kurki MI, Moonshine AL, Natarajan P, Orozco L, Peloso GM, Poplin R, Rivas MA, Ruano-Rubio V, Rose SA, Ruderfer DM, Shakir K, Stenson PD, Stevens C, Thomas BP, Tiao G, Tusie-Luna MT, Weisburd B, Won H-H, Yu D, Altshuler DM, Ardissino D, Boehnke M, Danesh J, Donnelly S, Elosua R, Florez JC, Gabriel SB, Getz G, Glatt SJ, Hultman CM, Kathiresan S, Laakso M, McCarroll S, McCarthy MI, McGovern D, McPherson R, Neale BM, Palotie A, Purcell SM, Saleheen D, Scharf JM, Sklar P, Sullivan PF, Tuomilehto J, Tsuang MT, Watkins HC, Wilson JG, Daly MJ, MacArthur DG, Exome Aggregation C. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016;536(7616):285-91.

148. Singh T, Walters JTR, Johnstone M, Curtis D, Suvisaari J, Torniainen M, Rees E, Iyegbe C, Blackwood D, McIntosh AM, Kirov G, Geschwind D, Murray RM, Di Forti M, Bramon E, Gandal M, Hultman CM, Sklar P, Study I, Consortium UK, Palotie A, Sullivan PF, O'Donovan MC, Owen MJ, Barrett JC. The contribution of rare variants to risk of schizophrenia in individuals with and without intellectual disability. Nat Genet. 2017;49(8):1167-73.

149. Vaaralahti K, Laitinen E-M, Tommiska J, Raivio T, Wehkalampi K, Dunkel L. LIN28B in Constitutional Delay of Growth and Puberty. The Journal of Clinical Endocrinology & Metabolism. 2010;95(6):3063-6.

150. Milano A, McMahon A, Welter D, Bowler E, Hastings E, Cunningham F, MacArthur J, Morales J, Gil L, Cerezo M, Flicek P, Burdett T, Pendlington ZM, Parkinson H, Junkins H, Hindorff L, Hall P. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). Nucleic Acids Res. 2016;45(D1):D896-D901.

151. Spain SL, Barrett JC. Strategies for fine-mapping complex traits. Hum Mol Genet. 2015;24(R1):R111-9.

152. Benner C, Spencer CCA, Havulinna AS, Salomaa V, Ripatti S, Pirinen M. FINEMAP: efficient variable selection using summary data from genome-wide association studies. Bioinformatics. 2016;32(10):1493-501.

153. Smemo S, Tena JJ, Kim K-H, Gamazon ER, Sakabe NJ, Gómez-Marín C, Aneas I, Credidio FL, Sobreira DR, Wasserman NF, Lee JH, Puviindran V, Tam D, Shen M, Son JE, Vakili NA, Sung H-K, Naranjo S, Acemel RD,

Manzanares M, Nagy A, Cox NJ, Hui C-C, Gomez-Skarmeta JL, Nóbrega MA. Obesity-associated variants within FTO form long-range functional connections with IRX3. Nature. 2014;507:371.
154. Claussnitzer M, Dankel SN, Kim K-H, Quon G, Meuleman W, Haugen C, Glunk V, Sousa IS, Beaudry JL, Puviindran V, Abdennur NA, Liu J, Svensson P-A, Hsu Y-H, Drucker DJ, Mellgren G, Hui C-C, Hauner H, Kellis M. FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. N Engl J Med. 2015;373(10):895-907.
155. Visscher PM, Goddard ME. From R.A. Fisher's 1918 Paper to GWAS a Century Later. 2019;211(4):1125-30.
156. Fisher RA. XV.—The Correlation between Relatives on the Supposition of Mendelian Inheritance. Transactions of the Royal Society of Edinburgh. 1919;52(2):399-433.
157. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TFC, McCarroll SA, Visscher PM. Finding the missing heritability of complex diseases. Nature. 2009;461:747.
158. Bush WS, Moore JH. Chapter 11: Genome-wide association studies. PLoS Comput Biol. 2012;8(12):e1002822-e.
159. Boyle EA, Li YI, Pritchard JK. An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell. 2017;169(7):1177-86.
160. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW, Goddard ME, Visscher PM. Common SNPs explain a large proportion of the heritability for human height. Nat Genet. 2010;42:565.
161. Shi H, Kichaev G, Pasaniuc B. Contrasting the Genetic Architecture of 30 Complex Traits from Summary Association Data. The American Journal of Human Genetics. 2016;99(1):139-53.
162. Marouli E, Graff M, Medina-Gomez C, Lo KS, Wood AR, Kjaer TR, Fine RS, Lu Y, Schurmann C, Highland HM, Rüeger S, Thorleifsson G, Justice AE, Lamparter D, Stirrups KE, Turcot V, Young KL, Winkler TW, Esko T, Karaderi T, Locke AE, Masca NGD, Ng MCY, Mudgal P, Rivas MA, Vedantam S, Mahajan A, Guo X, Abecasis G, Aben KK, Adair LS, Alam DS, Albrecht E, Allin KH, Allison M, Amouyel P, Appel EV, Arveiler D, Asselbergs FW, Auer PL, Balkau B, Banas B, Bang LE, Benn M, Bergmann S, Bielak LF, Blüher M, Boeing H, Boerwinkle E, Böger CA, Bonnycastle LL, Bork-Jensen J, Bots ML, Bottinger EP, Bowden DW, Brandslund I, Breen G, Brilliant MH, Broer L, Burt AA, Butterworth AS, Carey DJ, Caulfield MJ, Chambers JC, Chasman DI, Chen Y-DI, Chowdhury R, Christensen C, Chu AY, Cocca M, Collins FS, Cook JP, Corley J, Galbany JC, Cox AJ, Cuellar-Partida G, Danesh J, Davies G, de Bakker PIW, de Borst GJ, de Denus S, de Groot MCH, de Mutsert R, Deary IJ, Dedoussis G, Demerath EW, den Hollander AI, Dennis JG, Di Angelantonio E, Drenos F, Du M, Dunning AM, Easton DF, Ebeling T, Edwards TL, Ellinor PT, Elliott P, Evangelou E, Farmaki A-E, Faul JD, Feitosa MF, Feng S, Ferrannini E, Ferrario MM, Ferrieres J, Florez JC, Ford I, Fornage M, Franks PW, Frikke-Schmidt R, Galesloot TE, Gan W, Gandin I, Gasparini P, Giedraitis V, Giri A, Girotto G, Gordon SD, Gordon-Larsen P, Gorski M, Grarup N, Grove ML, Gudnason V, Gustafsson S, Hansen T, Harris KM, Harris TB, Hattersley AT, Hayward C, He L, Heid IM, Heikkilä K, Helgeland Ø, Hernesniemi J, Hewitt AW, Hocking LJ, Hollensted M, Holmen OL, Hovingh GK, Howson JMM, Hoyng CB, Huang PL, Hveem K, Ikram MA, Ingelsson E, Jackson AU, Jansson J-H, Jarvik GP, Jensen GB, Jhun MA, Jia Y, Jiang X, Johansson S, Jørgensen ME, Jørgensen T, Jousilahti P, Jukema JW, Kahali B, Kahn RS, Kähönen M, Kamstrup PR, Kanoni S, Kaprio J, Karaleftheri M, Kardia SLR, Karpe F, Kee F, Keeman R, Kiemeney LA, Kitajima H, Kluivers KB, Kocher T, Komulainen P, Kontto J, Kooner JS, Kooperberg C, Kovacs P, Kriebel J, Kuivaniemi H, Küry S, Kuusisto J, La Bianca M, Laakso M, Lakka TA, Lange EM, Lange LA, Langefeld CD, Langenberg C, Larson EB, Lee IT, Lehtimäki T, Lewis CE, Li H, Li J, Li-Gao R, Lin H, Lin L-A, Lin X, Lind L, Lindström J, Linneberg A, Liu Y, Liu Y, Lophatananon A, Luan Ja, Lubitz SA, Lyytikäinen L-P, Mackey DA, Madden PAF, Manning AK, Männistö S, Marenne G, Marten J, Martin NG, Mazul AL, Meidtner K, Metspalu A, Mitchell P, Mohlke KL, Mook-Kanamori DO, Morgan A, Morris AD, Morris AP, Müller-Nurasyid M, Munroe PB, Nalls MA, Nauck M, Nelson CP, Neville M, Nielsen SF, Nikus K, Njølstad PR, Nordestgaard BG, Ntalla I, O'Connel JR, Oksa H, Loohuis LMO, Ophoff RA, Owen KR, Packard CJ, Padmanabhan S, Palmer CNA, Pasterkamp G, Patel AP, Pattie A, Pedersen O, Peissig PL, Peloso GM, Pennell CE, Perola M, Perry JA, Perry JRB, Person TN, Pirie A, Polasek O, Posthuma D, Raitakari OT, Rasheed A, Rauramaa R, Reilly DF, Reiner AP, Renström F, Ridker PM, Rioux JD, Robertson N, Robino A, Rolandsson O, Rudan I, Ruth KS, Saleheen D, Salomaa V, Samani NJ, Sandow K, Sapkota Y, Sattar N, Schmidt MK, Schreiner PJ, Schulze MB, Scott RA, Segura-Lepe MP, Shah S, Sim X, Sivapalaratnam S, Small KS, Smith AV, Smith JA, Southam L, Spector TD, Speliotes EK, Starr JM, Steinthorsdottir V, Stringham HM, Stumvoll M, Surendran P, 't Hart LM, Tansey KE, Tardif J-C, Taylor KD, Teumer A, Thompson DJ, Thorsteinsdottir U, Thuesen BH, Tönjes A, Tromp G, Trompet S, Tsafantakis E, Tuomilehto J, Tybjaerg-Hansen A, Tyrer JP, Uher R, Uitterlinden AG, Ulivi S, van der Laan SW, Van Der Leij AR, van Duijn CM, van Schoor NM, van Setten J, Varbo A, Varga TV, Varma R, Edwards DRV, Vermeulen SH, Vestergaard H, Vitart V, Vogt TF, Vozzi D, Walker M, Wang F, Wang CA, Wang S, Wang Y, Wareham NJ, Warren HR, Wessel J, Willems SM, Wilson JG, Witte DR, Woods MO, Wu Y, Yaghootkar H, Yao J, Yao P, Yerges-Armstrong LM, Young R, Zeggini E, Zhan X, Zhang W, Zhao JH, Zhao W, Zhao W, Zheng H, Zhou W, The E-IC, Consortium CHDE, Exome BPC, Consortium TD-G, Go TDGC, Global Lipids Genetics C, ReproGen C,

Investigators M, Rotter JI, Boehnke M, Kathiresan S, McCarthy MI, Willer CJ, Stefansson K, Borecki IB, Liu DJ, North KE, Heard-Costa NL, Pers TH, Lindgren CM, Oxvig C, Kutalik Z, Rivadeneira F, Loos RJF, Frayling TM, Hirschhorn JN, Deloukas P, Lettre G. Rare and low-frequency coding variants alter human adult height. Nature. 2017;542:186.
163. Simons YB, Turchin MC, Pritchard JK, Sella G. The deleterious mutation load is insensitive to recent population history. Nat Genet. 2014;46(3):220-4.
164. De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Ercument Cicek A, Kou Y, Liu L, Fromer M, Walker S, Singh T, Klei L, Kosmicki J, Fu S-C, Aleksic B, Biscaldi M, Bolton PF, Brownfeld JM, Cai J, Campbell NG, Carracedo A, Chahrour MH, Chiocchetti AG, Coon H, Crawford EL, Crooks L, Curran SR, Dawson G, Duketis E, Fernandez BA, Gallagher L, Geller E, Guter SJ, Sean Hill R, Ionita-Laza I, Jimenez Gonzalez P, Kilpinen H, Klauck SM, Kolevzon A, Lee I, Lei J, Lehtimäki T, Lin C-F, Ma'ayan A, Marshall CR, McInnes AL, Neale B, Owen MJ, Ozaki N, Parellada M, Parr JR, Purcell S, Puura K, Rajagopalan D, Rehnström K, Reichenberg A, Sabo A, Sachse M, Sanders SJ, Schafer C, Schulte-Rüther M, Skuse D, Stevens C, Szatmari P, Tammimies K, Valladares O, Voran A, Wang L-S, Weiss LA, Jeremy Willsey A, Yu TW, Yuen RKC, The DDDS, Homozygosity Mapping Collaborative for A, Consortium UK, The Autism Sequencing C, Cook EH, Freitag CM, Gill M, Hultman CM, Lehner T, Palotie A, Schellenberg GD, Sklar P, State MW, Sutcliffe JS, Walsh CA, Scherer SW, Zwick ME, Barrett JC, Cutler DJ, Roeder K, Devlin B, Daly MJ, Buxbaum JD. Synaptic, transcriptional and chromatin genes disrupted in autism. Nature. 2014;515:209.
165. Fromer M, Pocklington AJ, Kavanagh DH, Williams HJ, Dwyer S, Gormley P, Georgieva L, Rees E, Palta P, Ruderfer DM, Carrera N, Humphreys I, Johnson JS, Roussos P, Barker DD, Banks E, Milanova V, Grant SG, Hannon E, Rose SA, Chambert K, Mahajan M, Scolnick EM, Moran JL, Kirov G, Palotie A, McCarroll SA, Holmans P, Sklar P, Owen MJ, Purcell SM, O'Donovan MC. De novo mutations in schizophrenia implicate synaptic networks. Nature. 2014;506:179.
166. Purcell SM, Moran JL, Fromer M, Ruderfer D, Solovieff N, Roussos P, O'Dushlaine C, Chambert K, Bergen SE, Kähler A, Duncan L, Stahl E, Genovese G, Fernández E, Collins MO, Komiyama NH, Choudhary JS, Magnusson PKE, Banks E, Shakir K, Garimella K, Fennell T, DePristo M, Grant SGN, Haggarty SJ, Gabriel S, Scolnick EM, Lander ES, Hultman CM, Sullivan PF, McCarroll SA, Sklar P. A polygenic burden of rare disruptive mutations in schizophrenia. Nature. 2014;506:185.
167. Lupski JR, Belmont JW, Boerwinkle E, Gibbs RA. Clan genomics and the complex architecture of human disease. Cell. 2011;147(1):32-43.
168. Wainschtein P, Jain DP, Yengo L, Zheng Z, Cupples LA, Shadyab AH, McKnight B, Shoemaker BM, Mitchell BD, Psaty BM, Kooperberg C, Roden D, Darbar D, Arnett DK, Regan EA, Boerwinkle E, Rotter JI, Allison MA, McDonald M-LN, Chung MK, Smith NL, Ellinor PT, Vasan RS, Mathias RA, Rich SS, Heckbert SR, Redline S, Guo X, Chen Y-DI, Liu C-T, de Andrade M, Yanek LR, Albert CM, Hernandez RD, McGarvey ST, North KE, Lange LA, Weir BS, Laurie CC, Yang J, Visscher PM. Recovery of trait heritability from whole genome sequence data. 2019:588020.
169. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. Science. 2015;348(6235):648.
170. Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ. Trait-Associated SNPs Are More Likely to Be eQTLs: Annotation to Enhance Discovery from GWAS. PLOS Genetics. 2010;6(4):e1000888.
171. Denny JC, Bastarache L, Roden DM. Phenome-Wide Association Studies as a Tool to Advance Precision Medicine. 2016;17(1):353-73.
172. Grunwald DJ, Eisen JS. Headwaters of the zebrafish — emergence of a new model vertebrate. Nature Reviews Genetics. 2002;3(9):717-24.
173. Meyers JR. Zebrafish: Development of a Vertebrate Model Organism. Current Protocols. 2018;16(1):e19.
174. Rafferty SA, Quinn TA. A beginner's guide to understanding and implementing the genetic modification of zebrafish. Prog Biophys Mol Biol. 2018;138:3-19.
175. Howe K, Clark MD, Torroja CF, Torrance J, Berthelot C, Muffato M, Collins JE, Humphray S, McLaren K, Matthews L, McLaren S, Sealy I, Caccamo M, Churcher C, Scott C, Barrett JC, Koch R, Rauch G-J, White S, Chow W, Kilian B, Quintais LT, Guerra-Assunção JA, Zhou Y, Gu Y, Yen J, Vogel J-H, Eyre T, Redmond S, Banerjee R, Chi J, Fu B, Langley E, Maguire SF, Laird GK, Lloyd D, Kenyon E, Donaldson S, Sehra H, Almeida-King J, Loveland J, Trevanion S, Jones M, Quail M, Willey D, Hunt A, Burton J, Sims S, McLay K, Plumb B, Davis J, Clee C, Oliver K, Clark R, Riddle C, Elliott D, Threadgold G, Harden G, Ware D, Begum S, Mortimore B, Kerry G, Heath P, Phillimore B, Tracey A, Corby N, Dunn M, Johnson C, Wood J, Clark S, Pelan S, Griffiths G, Smith M, Glithero R, Howden P, Barker N, Lloyd C, Stevens C, Harley J, Holt K, Panagiotidis G, Lovell J, Beasley H, Henderson C, Gordon D, Auger K, Wright D, Collins J, Raisen C, Dyer L, Leung K, Robertson L, Ambridge K, Leongamornlert D, McGuire S, Gilderthorp R, Griffiths C, Manthravadi D, Nichol S, Barker G, Whitehead S, Kay M, Brown J, Murnane C, Gray E, Humphries M, Sycamore N, Barker D, Saunders D, Wallis J, Babbage A, Hammond S, Mashreghi-Mohammadi M, Barr L, Martin S, Wray P, Ellington A, Matthews N, Ellwood M, Woodmansey R, Clark G, Cooper JD, Tromans A, Grafham D, Skuce C, Pandian R, Andrews R, Harrison E, Kimberley A, Garnett J, Fosker N, Hall R, Garner P, Kelly D, Bird C, Palmer S, Gehring I, Berger A, Dooley CM, Ersan-Ürün Z, Eser C, Geiger H, Geisler M, Karotki L, Kirn A, Konantz J, Konantz M, Oberländer M, Rudolph-Geiger S, Teucke M, Lanz C, Raddatz G, Osoegawa K, Zhu B, Rapp A, Widaa S, Langford C, Yang F,

Schuster SC, Carter NP, Harrow J, Ning Z, Herrero J, Searle SMJ, Enright A, Geisler R, Plasterk RHA, Lee C, Westerfield M, de Jong PJ, Zon LI, Postlethwait JH, Nüsslein-Volhard C, Hubbard TJP, Crollius HR, Rogers J, Stemple DL. The zebrafish reference genome sequence and its relationship to the human genome. Nature. 2013;496:498.
176. Wilson CA, High SK, McCluskey BM, Amores A, Yan Y-l, Titus TA, Anderson JL, Batzel P, Carvan MJ, Schartl M, Postlethwait JH. Wild Sex in Zebrafish: Loss of the Natural Sex Determinant in Domesticated Strains. Genetics. 2014;198(3):1291-308.
177. Gemberling M, Bailey TJ, Hyde DR, Poss KD. The zebrafish as a model for complex tissue regeneration. Trends in genetics : TIG. 2013;29(11):611-20.
178. Tang H, Liu Y, Luo D, Ogawa S, Yin Y, Li S, Zhang Y, Hu W, Parhar IS, Lin H, Liu X, Cheng CHK. The kiss/kissr systems are dispensable for zebrafish reproduction: evidence from gene knockout studies. Endocrinology. 2015;156(2):589-99.
179. Marvel M, Zmora N, Spicer OS, Wong T-T, Zohar Y. Knockout of the Gnrh genes in zebrafish: effects on reproduction and potential compensation by reproductive and feeding-related neuropeptides†. Biol Reprod. 2018;99(3):565-77.
180. Burden N, Chapman K, Sewell F, Robinson V. Pioneering better science through the 3Rs: an introduction to the national centre for the replacement, refinement, and reduction of animals in research (NC3Rs). Journal of the American Association for Laboratory Animal Science : JAALAS. 2015;54(2):198-208.
181. Goodman J, Chandna A, Roe K. Trends in animal use at US research facilities. J Med Ethics. 2015;41(7):567-9.
182. Lawson Nathan D, Wolfe Scot A. Forward and Reverse Genetic Approaches for the Analysis of Vertebrate Development in the Zebrafish. Dev Cell. 2011;21(1):48-64.
183. Li M, Zhao L, Page-McCaw PS, Chen W. Zebrafish Genome Engineering Using the CRISPR–Cas9 System. Trends Genet. 2016;32(12):815-27.
184. Blum M, De Robertis Edward M, Wallingford John B, Niehrs C. Morpholinos: Antisense and Sensibility. Dev Cell. 2015;35(2):145-9.
185. Kok Fatma O, Shin M, Ni C-W, Gupta A, Grosse Ann S, van Impel A, Kirchmaier Bettina C, Peterson-Maduro J, Kourkoulis G, Male I, DeSantis Dana F, Sheppard-Tindell S, Ebarasi L, Betsholtz C, Schulte-Merker S, Wolfe Scot A, Lawson Nathan D. Reverse Genetic Screening Reveals Poor Correlation between Morpholino-Induced and Mutant Phenotypes in Zebrafish. Dev Cell. 2015;32(1):97-108.
186. Rosen JN, Sweeney MF, Mably JD. Microinjection of zebrafish embryos to analyze gene function. Journal of visualized experiments : JoVE. 2009(25):1115.
187. Doudna JA, Charpentier E. The new frontier of genome engineering with CRISPR-Cas9. Science. 2014;346(6213):1258096.
188. Juolevi A, Sundvall J, Valsta L, Peltonen M, Jousilahti P, Puska P, Männistö S, Laatikainen T, Salomaa V, Vartiainen E. Thirty-five-year trends in cardiovascular risk factors in Finland. Int J Epidemiol. 2009;39(2):504-18.
189. Kaslin J, Panula P. Comparative anatomy of the histaminergic and other aminergic systems in zebrafish (Danio rerio). J Comp Neurol. 2001;440(4):342-77.
190. Kaslin J, Nystedt JM, Östergård M, Peitsaro N, Panula P. The Orexin/Hypocretin System in Zebrafish Is Connected to the Aminergic and Cholinergic Systems. J Neurosci. 2004;24(11):2678-89.
191. Kallio MA, Tuimala JT, Hupponen T, Klemelä P, Gentile M, Scheinin I, Koski M, Käki J, Korpelainen EI. Chipster: user-friendly analysis software for microarray and other high-throughput data. BMC Genomics. 2011;12(1):507.
192. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics (Oxford, England). 2010;26(1):139-40.
193. Team RC. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2017.
194. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics. 2009;10(1):48.
195. Supek F, Bošnjak M, Škunca N, Šmuc T. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. PLoS One. 2011;6(7):e21800.
196. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. Science. 2015;348(6235):648-60.
197. Thisse C, Thisse B. High-resolution in situ hybridization to whole-mount zebrafish embryos. Nat Protoc. 2007;3:59.
198. Loh P-R, Tucker G, Bulik-Sullivan BK, Vilhjálmsson BJ, Finucane HK, Salem RM, Chasman DI, Ridker PM, Neale BM, Berger B, Patterson N, Price AL. Efficient Bayesian mixed-model analysis increases association power in large cohorts. Nat Genet. 2015;47:284.
199. Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, Chu AY, Estrada K, Luan Ja, Kutalik Z, Amin N, Buchkovich ML, Croteau-Chonka DC, Day FR, Duan Y, Fall T, Fehrmann R, Ferreira T, Jackson AU, Karjalainen J, Lo

KS, Locke AE, Mägi R, Mihailov E, Porcu E, Randall JC, Scherag A, Vinkhuyzen AAE, Westra H-J, Winkler TW, Workalemahu T, Zhao JH, Absher D, Albrecht E, Anderson D, Baron J, Beekman M, Demirkan A, Ehret GB, Feenstra B, Feitosa MF, Fischer K, Fraser RM, Goel A, Gong J, Justice AE, Kanoni S, Kleber ME, Kristiansson K, Lim U, Lotay V, Lui JC, Mangino M, Leach IM, Medina-Gomez C, Nalls MA, Nyholt DR, Palmer CD, Pasko D, Pechlivanis S, Prokopenko I, Ried JS, Ripke S, Shungin D, Stancáková A, Strawbridge RJ, Sung YJ, Tanaka T, Teumer A, Trompet S, van der Laan SW, van Setten J, Van Vliet-Ostaptchouk JV, Wang Z, Yengo L, Zhang W, Afzal U, Ärnlöv J, Arscott GM, Bandinelli S, Barrett A, Bellis C, Bennett AJ, Berne C, Blüher M, Bolton JL, Böttcher Y, Boyd HA, Bruinenberg M, Buckley BM, Buyske S, Caspersen IH, Chines PS, Clarke R, Claudi-Boehm S, Cooper M, Daw EW, De Jong PA, Deelen J, Delgado G, Denny JC, Dhonukshe-Rutten R, Dimitriou M, Doney ASF, Dörr M, Eklund N, Eury E, Folkersen L, Garcia ME, Geller F, Giedraitis V, Go AS, Grallert H, Grammer TB, Gräßler J, Grönberg H, de Groot LCPGM, Groves CJ, Haessler J, Hall P, Haller T, Hallmans G, Hannemann A, Hartman CA, Hassinen M, Hayward C, Heard-Costa NL, Helmer Q, Hemani G, Henders AK, Hillege HL, Hlatky MA, Hoffmann W, Hoffmann P, Holmen O, Houwing-Duistermaat JJ, Illig T, Isaacs A, James AL, Jeff J, Johansen B, Johansson Å, Jolley J, Juliusdottir T, Junttila J, Kho AN, Kinnunen L, Klopp N, Kocher T, Kratzer W, Lichtner P, Lind L, Lindström J, Lobbens S, Lorentzon M, Lu Y, Lyssenko V, Magnusson PKE, Mahajan A, Maillard M, McArdle WL, McKenzie CA, McLachlan S, McLaren PJ, Menni C, Merger S, Milani L, Moayyeri A, Monda KL, Morken MA, Müller G, Müller-Nurasyid M, Musk AW, Narisu N, Nauck M, Nolte IM, Nöthen MM, Oozageer L, Pilz S, Rayner NW, Renstrom F, Robertson NR, Rose LM, Roussel R, Sanna S, Scharnagl H, Scholtens S, Schumacher FR, Schunkert H, Scott RA, Sehmi J, Seufferlein T, Shi J, Silventoinen K, Smit JH, Smith AV, Smolonska J, Stanton AV, Stirrups K, Stott DJ, Stringham HM, Sundström J, Swertz MA, Syvänen A-C, Tayo BO, Thorleifsson G, Tyrer JP, van Dijk S, van Schoor NM, van der Velde N, van Heemst D, van Oort FVA, Vermeulen SH, Verweij N, Vonk JM, Waite LL, Waldenberger M, Wennauer R, Wilkens LR, Willenborg C, Wilsgaard T, Wojczynski MK, Wong A, Wright AF, Zhang Q, Arveiler D, Bakker SJL, Beilby J, Bergman RN, Bergmann S, Biffar R, Blangero J, Boomsma DI, Bornstein SR, Bovet P, Brambilla P, Brown MJ, Campbell H, Caulfield MJ, Chakravarti A, Collins R, Collins FS, Crawford DC, Cupples LA, Danesh J, de Faire U, den Ruijter HM, Erbel R, Erdmann J, Eriksson JG, Farrall M, Ferrannini E, Ferrières J, Ford I, Forouhi NG, Forrester T, Gansevoort RT, Gejman PV, Gieger C, Golay A, Gottesman O, Gudnason V, Gyllensten U, Haas DW, Hall AS, Harris TB, Hattersley AT, Heath AC, Hengstenberg C, Hicks AA, Hindorff LA, Hingorani AD, Hofman A, Hovingh GK, Humphries SE, Hunt SC, Hypponen E, Jacobs KB, Jarvelin M-R, Jousilahti P, Jula AM, Kaprio J, Kastelein JJP, Kayser M, Kee F, Keinanen-Kiukaanniemi SM, Kiemeney LA, Kooner JS, Kooperberg C, Koskinen S, Kovacs P, Kraja AT, Kumari M, Kuusisto J, Lakka TA, Langenberg C, Le Marchand L, Lehtimäki T, Lupoli S, Madden PAF, Männistö S, Manunta P, Marette A, Matise TC, McKnight B, Meitinger T, Moll FL, Montgomery GW, Morris AD, Morris AP, Murray JC, Nelis M, Ohlsson C, Oldehinkel AJ, Ong KK, Ouwehand WH, Pasterkamp G, Peters A, Pramstaller PP, Price JF, Qi L, Raitakari OT, Rankinen T, Rao DC, Rice TK, Ritchie M, Rudan I, Salomaa V, Samani NJ, Saramies J, Sarzynski MA, Schwarz PEH, Sebert S, Sever P, Shuldiner AR, Sinisalo J, Steinthorsdottir V, Stolk RP, Tardif J-C, Tönjes A, Tremblay A, Tremoli E, Virtamo J, Vohl M-C, The Electronic Medical R, Genomics C, The MC, The PC, The LifeLines Cohort S, Amouyel P, Asselbergs FW, Assimes TL, Bochud M, Boehm BO, Boerwinkle E, Bottinger EP, Bouchard C, Cauchi S, Chambers JC, Chanock SJ, Cooper RS, de Bakker PIW, Dedoussis G, Ferrucci L, Franks PW, Froguel P, Groop LC, Haiman CA, Hamsten A, Hayes MG, Hui J, Hunter DJ, Hveem K, Jukema JW, Kaplan RC, Kivimaki M, Kuh D, Laakso M, Liu Y, Martin NG, März W, Melbye M, Moebus S, Munroe PB, Njølstad I, Oostra BA, Palmer CNA, Pedersen NL, Perola M, Pérusse L, Peters U, Powell JE, Power C, Quertermous T, Rauramaa R, Reinmaa E, Ridker PM, Rivadeneira F, Rotter JI, Saaristo TE, Saleheen D, Schlessinger D, Slagboom PE, Snieder H, Spector TD, Strauch K, Stumvoll M, Tuomilehto J, Uusitupa M, van der Harst P, Völzke H, Walker M, Wareham NJ, Watkins H, Wichmann HE, Wilson JF, Zanen P, Deloukas P, Heid IM, Lindgren CM, Mohlke KL, Speliotes EK, Thorsteinsdottir U, Barroso I, Fox CS, North KE, Strachan DP, Beckmann JS, Berndt SI, Boehnke M, Borecki IB, McCarthy MI, Metspalu A, Stefansson K, Uitterlinden AG, van Duijn CM, Franke L, Willer CJ, Price AL, Lettre G, Loos RJF, Weedon MN, Ingelsson E, O'Connell JR, Abecasis GR, Chasman DI, Goddard ME, Visscher PM, Hirschhorn JN, Frayling TM. Defining the role of common variation in the genomic and biological architecture of adult human height. Nat Genet. 2014;46:1173.

200. Justice AE, Winkler TW, Feitosa MF, Graff M, Fisher VA, Young K, Barata L, Deng X, Czajkowski J, Hadley D, Ngwa JS, Ahluwalia TS, Chu AY, Heard-Costa NL, Lim E, Perez J, Eicher JD, Kutalik Z, Xue L, Mahajan A, Renström F, Wu J, Qi Q, Ahmad S, Alfred T, Amin N, Bielak LF, Bonnefond A, Bragg J, Cadby G, Chittani M, Coggeshall S, Corre T, Direk N, Eriksson J, Fischer K, Gorski M, Neergaard Harder M, Horikoshi M, Huang T, Huffman JE, Jackson AU, Justesen JM, Kanoni S, Kinnunen L, Kleber ME, Komulainen P, Kumari M, Lim U, Luan Ja, Lyytikäinen L-P, Mangino M, Manichaikul A, Marten J, Middelberg RPS, Müller-Nurasyid M, Navarro P, Pérusse L, Pervjakova N, Sarti C, Smith AV, Smith JA, Stančáková A, Strawbridge RJ, Stringham HM, Sung YJ, Tanaka T, Teumer A, Trompet S, van der Laan SW, van der Most PJ, Van Vliet-Ostaptchouk JV, Vedantam SL, Verweij N, Vink JM, Vitart V, Wu Y, Yengo L, Zhang W, Hua Zhao J, Zimmermann ME, Zubair N, Abecasis GR, Adair LS, Afaq S, Afzal U, Bakker SJL, Bartz TM, Beilby J, Bergman RN, Bergmann S, Biffar R, Blangero J, Boerwinkle E, Bonnycastle LL, Bottinger E, Braga D, Buckley BM, Buyske S, Campbell H, Chambers JC, Collins FS, Curran JE, de Borst GJ, de Craen AJM, de Geus EJC, Dedoussis G, Delgado GE, den Ruijter HM, Eiriksdottir G, Eriksson AL, Esko T, Faul JD, Ford I, Forrester T, Gertow K, Gigante B,

Glorioso N, Gong J, Grallert H, Grammer TB, Grarup N, Haitjema S, Hallmans G, Hamsten A, Hansen T, Harris TB, Hartman CA, Hassinen M, Hastie ND, Heath AC, Hernandez D, Hindorff L, Hocking LJ, Hollensted M, Holmen OL, Homuth G, Jan Hottenga J, Huang J, Hung J, Hutri-Kähönen N, Ingelsson E, James AL, Jansson J-O, Jarvelin M-R, Jhun MA, Jørgensen ME, Juonala M, Kähönen M, Karlsson M, Koistinen HA, Kolcic I, Kolovou G, Kooperberg C, Krämer BK, Kuusisto J, Kvaløy K, Lakka TA, Langenberg C, Launer LJ, Leander K, Lee NR, Lind L, Lindgren CM, Linneberg A, Lobbens S, Loh M, Lorentzon M, Luben R, Lubke G, Ludolph-Donislawski A, Lupoli S, Madden PAF, Männikkö R, Marques-Vidal P, Martin NG, McKenzie CA, McKnight B, Mellström D, Menni C, Montgomery GW, Musk AW, Narisu N, Nauck M, Nolte IM, Oldehinkel AJ, Olden M, Ong KK, Padmanabhan S, Peyser PA, Pisinger C, Porteous DJ, Raitakari OT, Rankinen T, Rao DC, Rasmussen-Torvik LJ, Rawal R, Rice T, Ridker PM, Rose LM, Bien SA, Rudan I, Sanna S, Sarzynski MA, Sattar N, Savonen K, Schlessinger D, Scholtens S, Schurmann C, Scott RA, Sennblad B, Siemelink MA, Silbernagel G, Slagboom PE, Snieder H, Staessen JA, Stott DJ, Swertz MA, Swift AJ, Taylor KD, Tayo BO, Thorand B, Thuillier D, Tuomilehto J, Uitterlinden AG, Vandenput L, Vohl M-C, Völzke H, Vonk JM, Waeber G, Waldenberger M, Westendorp RGJ, Wild S, Willemsen G, Wolffenbuttel BHR, Wong A, Wright AF, Zhao W, Zillikens MC, Baldassarre D, Balkau B, Bandinelli S, Böger CA, Boomsma DI, Bouchard C, Bruinenberg M, Chasman DI, Chen Y-D, Chines PS, Cooper RS, Cucca F, Cusi D, Faire Ud, Ferrucci L, Franks PW, Froguel P, Gordon-Larsen P, Grabe H-J, Gudnason V, Haiman CA, Hayward C, Hveem K, Johnson AD, Wouter Jukema J, Kardia SLR, Kivimaki M, Kooner JS, Kuh D, Laakso M, Lehtimäki T, Marchand LL, März W, McCarthy MI, Metspalu A, Morris AP, Ohlsson C, Palmer LJ, Pasterkamp G, Pedersen O, Peters A, Peters U, Polasek O, Psaty BM, Qi L, Rauramaa R, Smith BH, Sørensen TIA, Strauch K, Tiemeier H, Tremoli E, van der Harst P, Vestergaard H, Vollenweider P, Wareham NJ, Weir DR, Whitfield JB, Wilson JF, Tyrrell J, Frayling TM, Barroso I, Boehnke M, Deloukas P, Fox CS, Hirschhorn JN, Hunter DJ, Spector TD, Strachan DP, van Duijn CM, Heid IM, Mohlke KL, Marchini J, Loos RJF, Kilpeläinen TO, Liu C-T, Borecki IB, North KE, Cupples LA. Genome-wide meta-analysis of 241,258 adults accounting for smoking behaviour identifies novel loci for obesity traits. Nature Communications. 2017;8:14977.

201. Yang D-H, Moss EG. Temporally regulated expression of Lin-28 in diverse tissues of the developing mouse. Gene Expression Patterns. 2003;3(6):719-26.

202. Wadhwa PD, Buss C, Entringer S, Swanson JM. Developmental Origins of Health and Disease: Brief History of the Approach and Current Focus on Epigenetic Mechanisms. Semin Reprod Med. 2009;27(5):358-68.

203. Leinonen JT, Chen Y-C, Tukiainen T, Panula P, Widén E. Transient modification of lin28b expression - permanent effects on zebrafish growth. Mol Cell Endocrinol. 2019;479:61-70.

204. Parichy DM, Elizondo MR, Mills MG, Gordon TN, Engeszer RE. Normal table of postembryonic zebrafish development: staging by externally visible anatomy of the living fish. Developmental dynamics : an official publication of the American Association of Anatomists. 2009;238(12):2975-3015.

205. Tanner JM, Whitehouse RH, Marubini E, Resele LF. The adolescent growth spurt of boys and girls of the Harpenden Growth Study. Ann Hum Biol. 1976;3(2):109-26.

206. El-Brolosy MA, Kontarakis Z, Rossi A, Kuenne C, Günther S, Fukuda N, Kikhi K, Boezio GLM, Takacs CM, Lai S-L, Fukuda R, Gerri C, Giraldez AJ, Stainier DYR. Genetic compensation triggered by mutant mRNA degradation. Nature. 2019.

207. Ogawa S, Kitahashi T, Parhar IS. Cloning and Expression of kiss2 in the Zebrafish and Medaka. Endocrinology. 2009;150(2):821-31.

208. Khacho M, Slack RS. Mitochondrial activity in the regulation of stem cell self-renewal and differentiation. Curr Opin Cell Biol. 2017;49:1-8.

209. Vega-Naredo I, Loureiro R, Mesquita KA, Barbosa IA, Tavares LC, Branco AF, Erickson JR, Holy J, Perkins EL, Carvalho RA, Oliveira PJ. Mitochondrial metabolism directs stemness and differentiation in P19 embryonal carcinoma stem cells. Cell Death Differ. 2014;21:1560.

210. Pickrell JK, Berisa T, Liu JZ, Ségurel L, Tung JY, Hinds DA. Detection and interpretation of shared genetic influences on 42 human traits. Nat Genet. 2016;48(7):709-17.

211. Mayer C, Acosta-Martinez M, Dubois SL, Wolfe A, Radovick S, Boehm U, Levine JE. Timing and completion of puberty in female mice depend on estrogen receptor α-signaling in kisspeptin neurons. Proc Natl Acad Sci U S A. 2010;107(52):22693-8.

212. Tiano JP, Nohara K, Waraich RS, Mauvais-Jarvis F, Laque A, Münzberg H, Zhang Y, Tong J. Early-Life Exposure to Testosterone Programs the Hypothalamic Melanocortin System. Endocrinology. 2011;152(4):1661-9.

213. Warrington NM, Shevroja E, Hemani G, Hysi PG, Jiang Y, Auton A, Boer CG, Mangino M, Wang CA, Kemp JP, McMahon G, Medina-Gomez C, Hickey M, Trajanoska K, Wolke D, Ikram MA, andMe Research T, Montgomery GW, Felix JF, Wright MJ, Mackey DA, Jaddoe VW, Martin NG, Tung JY, Davey Smith G, Pennell CE, Spector TD, van Meurs J, Rivadeneira F, Medland SE, Evans DM. Genome-wide association study identifies nine novel loci for 2D:4D finger ratio, a putative retrospective biomarker of testosterone exposure in utero. Hum Mol Genet. 2018;27(11):2025-38.

214. Snyder PJ, Peachey H, Hannoush P, Berlin JA, Loh L, Holmes JH, Dlewati A, Staley J, Santanna J, Kapoor SC, Attie MF, Haddad JG, Jr., Strom BL. Effect of Testosterone Treatment on Bone Mineral Density in Men Over 65 Years of Age1. The Journal of Clinical Endocrinology & Metabolism. 1999;84(6):1966-72.
215. García-Cruz E, Huguet J, Piqueras M, Ribal MJ, Alcaraz A. Second to fourth digit ratio, adult testosterone level and testosterone deficiency. 2012;109(2):266-71.

Recent Publications in this Series

50/2019 Veer Singh Marwah
Strategies to Improve Standardization and Robustness of Toxicogenomics Data Analysis
51/2019 Iryna Hlushchenko
Actin Regulation in Dendritic Spines: From Synaptic Plasticity to Animal Behavior and Human Neurodevelopmental Disorders
52/2019 Heini Liimatta
Effectiveness of Preventive Home Visits among Community-Dwelling Older People
53/2019 Helena Karppinen
Older People´s Views Related to Their End of Life: Will-to-Live, Wellbeing and Functioning
54/2019 Jenni Laitila
Elucidating Nebulin Expression and Function in Health and Disease
55/2019 Katarzyna Ciuba
Regulation of Contractile Actin Structures in Non-Muscle Cells
56/2019 Sami Blom
Spatial Characterisation of Prostate Cancer by Multiplex Immunohistochemistry and Quantitative Image Analysis
57/2019 Outi Lyytinen
Molecular Details of the Double-Stranded RNA Virus Replication and Assembly
58/2019 Markus Räsänen
Vascular Endothelial Growth Factor-B and the Bmx Tyrosine Kinase in Cardiac Hypertrophy and Revascularization
59/2019 Vuokko Nummi
Insights into Clinical and Laboratory Phenotypes of Von Willebrand Disease
60/2019 Shah Hasan
Challenges of Hyper-Prolificacy in the Pig: Colostrum and Gut Microbiota
61/2019 Sanna Matilainen
Pathomechanisms of Leigh Syndrome: Defects of Post-Transcriptional and Post-Translational Regulation of Mitochondrial Metabolism
62/2019 Kirsi Santti
Desmoid Tumor: Oncological Management and Prognostic Biomarkers
63/2019 Hesham E. Abdolhfid Mohamed
Evaluation of Prognostic Markers for Oropharyngeal Carcinoma Using Tissue Microarray
64/2019 Johanna Uhari-Väänänen
Contributions of μ- and κ-Opioidergic Systems to Ethanol Intake and Addiction
65/2019 Susanna Rapo-Pylkkö
Chronic Pain and Neuropathic Pain among Community-dwelling Older Adults in Primary Health Care Settings
66/2019 Helka Göös
Human Transcription Factor Protein-protein Interactions in Health and Disease
67/2019 Maiju Rinne
Molecular Evolution of G Protein-Coupled Receptors – Insights into the Orexin System
68/2019 Ester Orav
The Role of Kainate Receptor Auxiliary Subunits NETO1 and NETO2 in Development of Hippocampal Circuitry
69/2019 Liang Wang
Biological Functions of Novel Mitochondrial Proteins
70/2019 Timo Carpén
Novel Diagnostic and Prognostic Aspects of HPV-Related and -Unrelated Oropharyngeal Cancer

Helsinki 2019 ISSN 2342-3161 ISBN 978-951-51-5520-7